Estimation of cochlear frequency selectivity using a convolution model of forward-masked compound action potentials

François Deloche[1*], Satyabrata Parida[1,2], Andrew Sivaprakasam[2] and Michael G. Heinz[1,2]

[1*]Department of Speech, Language, and Hearing Sciences, Purdue University, Street, West Lafayette, 47907, Indiana, USA.
[2]Weldon School of Biomedical Engineering, Purdue University, Street, West Lafayette, 47907, Indiana, USA.

*Corresponding author(s). E-mail(s): francois.deloche@ugent.be;
Contributing authors: satyaparida@pitt.edu; asivapr@purdue.edu; mheinz@purdue.edu;

Abstract

Purpose: Frequency selectivity is a fundamental property of the peripheral auditory system; however, the invasiveness of auditory nerve (AN) experiments limits its study in the human ear. Compound action potentials (CAPs) associated with forward-masking have been suggested as an alternative to assess cochlear frequency selectivity. Previous methods relied on an empirical comparison of AN and CAP tuning curves in animal models, arguably not taking full advantage of the information contained in forward-masked CAP waveforms.
Methods: To improve the estimation of cochlear frequency selectivity based on the CAP, we introduce a convolution model to fit forward-masked CAP waveforms. The model generates masking patterns that, when convolved with a unitary response, can predict the masking of the CAP waveform induced by Gaussian noise maskers. Model parameters, including those characterizing frequency selectivity, are fine-tuned by minimizing waveform prediction errors across numerous masking conditions, yielding robust estimates.
Results: The method was applied to click-evoked CAPs at the round window of anesthetized chinchillas using notched-noise maskers

with various notch widths and attenuations. The estimated quality factor Q10 as a function of center frequency is shown to closely match the average quality factor obtained from AN fiber tuning curves, without the need for an empirical correction factor. **Conclusion:** This study establishes a moderately invasive method for estimating cochlear frequency selectivity with potential applicability to other animal species or humans. Beyond the estimation of frequency selectivity, the proposed model proved to be remarkably accurate in fitting forward-masked CAP responses, and could be extended to study more complex aspects of cochlear signal processing (e.g., compressive nonlinearities).

Keywords: compound action potential, auditory nerve, frequency selectivity, cochlear tuning, forward masking

1 Introduction

Much of our knowledge about the mammalian peripheral auditory system has been gained from single-fiber recordings of the auditory nerve in animals commonly used in laboratory studies. However, the invasiveness of these experiments prevents their use in humans, hindering the search of potential specificities of the human auditory system. Other means have been employed to infer the properties of the human inner ear, either through psychophysical experiments, or through less invasive physiological methods. In particular, a combination of these solutions – including psychophysical experiments based on masking [1, 2], otoacoustic emissions (OAEs) [1, 3] and compound action potentials (CAPs) [4] – has led to a growing body of evidence that cochlear frequency selectivity is sharper in humans than small mammals. Frequency selectivity is a fundamental property of the peripheral auditory system, but its study is not straightforward, a reason being that it is affected by cochlear compressive nonlinearities [2, 5, 6]. As a result, although data on cochlear frequency tuning in humans have been obtained by various means, the picture is not as detailed as for other mammals, and some methods of assessing cochlear frequency selectivity do not show any significant difference with small mammals [7]. To advance our knowledge in this area, it is necessary to refine the available tools and to better understand how they relate to auditory physiology. For example, OAE-based estimates of frequency selectivity would benefit from a better understanding of how OAE delays [8] or distortion-product level functions [9] relate to cochlear tuning. The focus of this paper is the CAP, an auditory evoked potential that reflects the summed activity of auditory nerve fibers (ANFs). CAP data can be obtained with a satisfactory signal-to-noise ratio (SNR) at the cost of moderate invasiveness [10, 11], and, if analyzed with an appropriate model, could provide significant information on the compound response of ANFs, including AN frequency tuning.

Estimation methods of cochlear frequency selectivity based on the CAP rely on the masking paradigm, similar to psychophysical experiments historically associated with the measurement of critical bands in humans [12, 13]. While simultaneous masking reflects both excitatory and suppressive masking [14–16], estimates based on forward masking reflect only excitatory masking and have good agreement with ANF tuning curves [15, 17, 18]. In the last decade, Verschooten et al. refined a previous estimation procedure based on forward-masked CAPs [15, 17] using notched-noise maskers. The procedure was first validated in animal models [18] and later applied to human subjects [4]. Their estimation method was based on establishing iso-response curves for masker level versus masker notch width. However, the method required an empirical correction to match the quality factor Q_{10} of ANF tuning curves with the correction factor varying across species. In particular, the estimate of Q_{10} for humans was higher if the correction factor found for macaques was applied instead of the factor found for smaller mammals, leaving the exact range for Q_{10} uncertain.

In this work, we attempt to reduce the dependence of the estimation of frequency selectivity on an empirical correction factor by relating forward-masked CAP responses to a computational model of ANF activity. To this end, we assume that the masked part of forward-masked CAP responses can be approximated by a 'masking pattern' defined in the time domain convolved with a unitary response. Convolution models have been used for decades to describe the CAP [19], but applications of these models have been limited since they require many assumptions about the factors affecting the CAP waveform. These factors include the (level-dependent) relationship between cochlear place and AN spike latencies, the spread of excitation along the cochlear partition, the spike unit response, and the distribution of thresholds and rate functions [20]. However, considering forward-masked CAPs with multiple masking stimuli, but with a fixed probe, simplifies the modeling approach because several factors remain constant as a consequence of using a unique and fixed probe. In addition, forward-masked CAPs provide information about some of the factors mentioned above, such as the place-latency relationship using high-pass noise maskers with different cut-off frequencies [10, 21]. In this paper, we introduce a model for predicting click-evoked CAP waveforms in the presence of notched-noise forward-maskers with different spectral properties. The estimation of the model parameters, including cochlear frequency selectivity, is done through the minimization of the waveform prediction errors. To assess our method, we recorded forward-masked CAPs at the round window of anesthetized chinchillas and tuned the model to fit the masked CAP waveforms. We found that the estimates of the quality factor averaged over experiments closely matched Q_{10} values from published ANF single-fiber data. Beyond the estimation of frequency selectivity, the results show that the model was remarkably accurate in fitting the forward-masked CAP responses, highlighting the potential of the proposed paradigm to study other properties of the peripheral auditory system (e.g., compressive nonlinearities).

2 Methods

2.1 Methods overview

Approach. The methods introduced in this paper build upon the convolution model that has been widely adopted since early work on the CAP [19, 20]. In this framework, in its most basic representation, a CAP waveform is written as a convolution between two components, a cochlear excitation pattern E and a unitary response u_0 that shares the biphasic shape of the CAP:

$$CAP\,(t) = E * u_0\,(t) = \int_\tau E(\tau)\, u_0(t-\tau)\,\mathrm{d}\tau\ . \tag{1}$$

The cochlear excitation pattern represents the distribution of excitation levels across different latencies. For consistency, we use the term "latency domain" throughout the paper to refer to the context before the convolution is applied, corresponding to the dummy variable τ in the above integral. Conversely, we use the term "time domain" to refer to the context after the convolution has been performed, when the excitation patterns and unitary response are combined to generate the CAP waveforms.

However, our focus in this work is on the part of the CAP that is affected by forward masking. In particular, we are interested in the differences of a click-evoked CAP waveform induced by spectral manipulations on a notched-noise masker (e.g., increasing the masker notch width). Rather than examining the raw CAP waveform, we consider $\Delta CAP(t)$, the release of masking of the CAP, defined by the difference in the CAP amplitude between two masking conditions:

$$\Delta CAP(t) = CAP_{\mathrm{masked}}(t) - CAP_{masked,b}(t)\,,$$

where $CAP_{\mathrm{masked}}(t)$ is the CAP response associated with a notched-noise masker and $CAP_{\mathrm{masked,\ b}}(t)$ is the response associated with a reference masking condition, chosen as the no-notch masker ('b' stands both for baseline or broadband noise).
We can write a similar equation to Eq 1 for ΔCAP, which will play a key role in the rest of the paper:

$$\Delta CAP(t) = R * u(t) \tag{2}$$

where we call $R(\tau)$ the *masking-release pattern* and u the unitary response.

The approach of this paper is to introduce a model that generates estimates for the masking-release pattern $R(\tau)$ associated with each presented forward-masking condition. The generation of these patterns depend on several parameters, including the quality factor Q_{10} characterizing cochlear frequency tuning. The model is fitted to experimental data by minimizing the mean squared error between the generated waveforms and the actual masking-release waveforms, resulting in estimates of the model parameters.

The structure of the Methods section is as follows. We start by introducing the concept of masking input-output curves, on which our model is based. The

model architecture is described next. The remaining subsections are dedicated to the application of the introduced methods to experimental data collected in anesthetized chinchillas. First, we describe the data collection procedure and provide more details on the masking conditions used. We then detail how we adjusted the model to the experimental data. In addition, Appendix A provides more context to the convolution model (Equations 1 and 2) with a closer examination of the underlying assumptions.

Stimulus paradigm As we proceed into the Methods section, this paragraph briefly describes the relevant experimental paradigm. We consider CAP waveforms evoked by a fixed probe (alternating-polarity click) in the presence of Gaussian noise maskers in a forward-masking setting. The level of the click probe is 80 dB peak-equivalent sound pressure level (peSPL). The CAP waveforms are obtained by averaging the responses (over the two polarities) associated with the same masker. Figure 1a shows the time representation of a stimulus cycle. The panel b of the same figure shows the spectral profile of the three types of maskers that were used in this study: high-pass noise maskers, notched-noise maskers with a varying notch amplitude, and notched-noise maskers with a varying notch width. The three types of maskers were designed to probe different aspects of the CAP, as further explained in the Methods section; all three types were needed to optimize the convolution model introduced next and to estimate the parameters of interest (in particular: the place-latency relationship, the growth of masking, and the cochlear frequency selectivity). The set of maskers also includes the reference condition, i.e., a broadband-noise masker without a spectral notch.

2.2 Model

This subsection describes the architecture of the model used to generate the estimates of the masking-release CAP waveforms. Its purpose is not to provide exhaustive information on how the model parameters were estimated, which is done at the end of the Methods section.

Masking input-output functions. To build our model, we make the assumption that the amount of masking of the CAP can be quantified using the outputs of a cochlear filter bank and masking input/output (I/O) curves determining the growth-of-masking for each output channel (Fig 2, steps B and C). More explicitly, if I is the average intensity in response to a masker at the output of a single cochlear filter, we assume that the amount of masking M for the compound response of the associated ANFs can be characterized by a function of I (masking I/O function). We used the Weibull cumulative distribution function (CDF), as in Verschooten et al. [18]:

$$M(I) = C\left[1 - e^{-\left(\frac{(I-I_0)_+}{\lambda}\right)^s}\right], \tag{3}$$

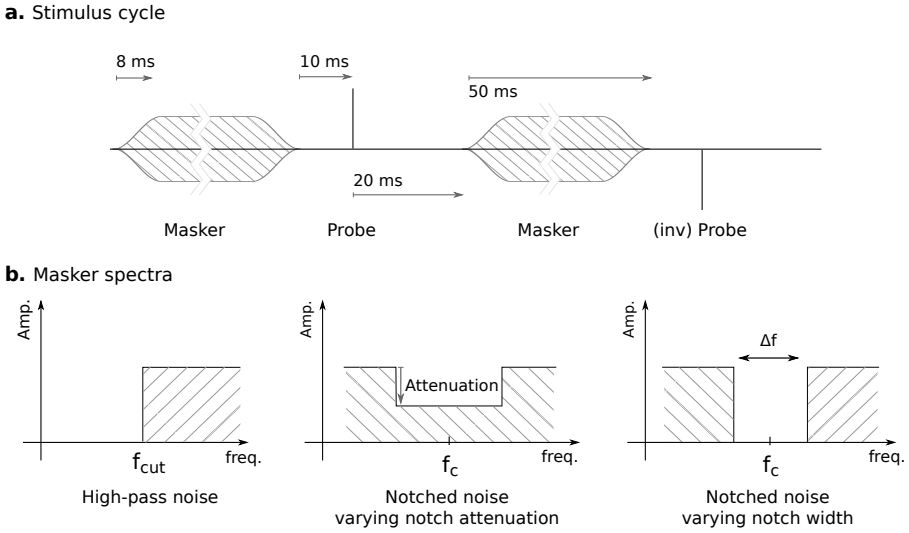

Fig. 1 Time representation of one stimulus cycle (**a.**) and spectral representation of the three types of maskers (**b.**). **a.** The stimuli consist of the repetition of a masker and probe. The masker is generated from Gaussian noise, following one of the frequency patterns represented in panel b. The probe is a 80-dB peSPL click of alternating polarity. The forward-masked CAPs are obtained by averaging the responses evoked by the probe presented under the same masking condition. The durations illustrated are from left to right: gating time (cosine ramp), masker-probe interval, probe-masker interval and masker duration. **b.** Schematic representation of the spectra of the three different types of maskers. Each type of masker was designed for a different purpose: high-pass noise maskers for the estimation of the place-latency relationship ('narrow-band analysis' method [21]), notched noise maskers for the estimation of masking input-output functions (maskers with varying notch attenuation) or frequency selectivity (varying notch widths).

where $(I - I_0)_+ = I - I_0$ if $I \geq I_0$, 0 elsewhere. Its parameters are I_0, λ (scale parameter), and s (shape parameter). The Weibull CDF is similar to a sigmoid, but does not impose symmetry around its half-maximum value point. By convention, we set the constant C so that the masking I/O functions are constrained to 100% masking for the response level.

Note that the above assumption is not the only basis of our model. In particular, other assumptions underlie the convolution model (Eq 2). To avoid overloading the Methods section, we leave the discussion of these additional assumptions in Appendix A.

Place-latency relationship. For our model to simulate forward-masked CAP waveforms, we need to have estimates of the relevant quantities (e.g., the amount of masking M or the masking-release pattern R) in the latency domain. However, it is generally easier to consider these values in the place (or center frequency, CF) domain. Whenever it is useful, we will convert dependencies on latencies into a dependency on CF (or vice versa) by assuming that CFs and latencies are related by a power-law: $CF(\tau) = B(\tau - t_0)^{\alpha}$ for $\tau > t_0$. In this equation, B, t_0 and α can be estimated using the high-pass noise maskers,

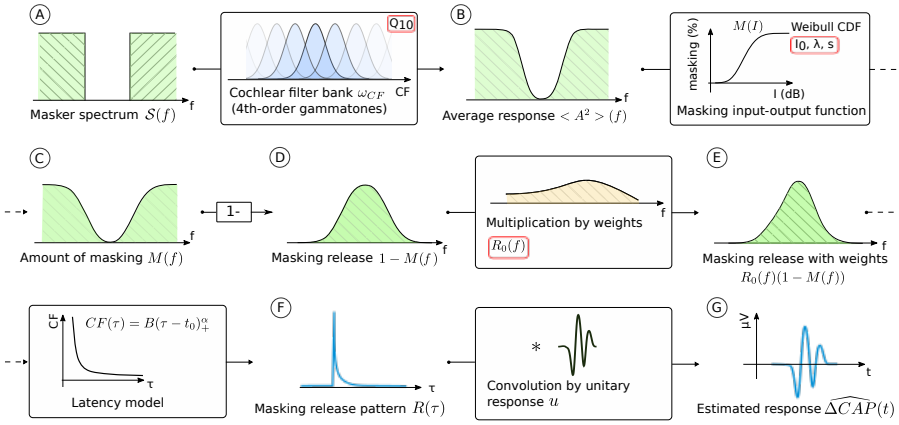

Fig. 2 Flow diagram of the generation of the masking-release estimates $\widehat{\Delta CAP}(t)$. The masker spectrum (**A**) is decomposed by a bank of gammatone filters. As the masker spectra are of simple form, i.e., composed of rectangular bands, the average response (**B**) at the output of the filter bank was computed using an analytical formula (see text). The masking input-output function applied to the average response provides the amount of masking $M(f)$ (**C**) or, equivalently, the amount of masking release $1 - M(f)$ (**D**). Frequency weights $R_0(f)$ are included to account for the non-homogeneous contributions of different CFs to ΔCAP. This yields the final estimate of the amount of masking release defined in the frequency domain (**E**). With a change of variable substituting CFs with latencies (using a power-law), the masking release is converted to the latency domain, giving the masking-release pattern $R(\tau)$ (**F**) . Once convolved with the unitary response u, we finally obtain the estimate of the release waveform $\widehat{\Delta CAP}(t)$ (**G**). The parameters that are fine-tuned during the optimization process (gradient descent) are highlighted in red: they are Q_{10}, the masking I/O function (Weibull CDF) variables, and the frequency weights. The unitary response u and the power-law parameters relating CFs and latencies are also parameters of the model, but are adjusted independently by a specific procedure (see text).

similar to the narrow-band analysis of the CAP already described in other studies [21, 22]. This method assumes that the high-pass noise stimuli mask the contributions of ANFs with CFs above the cut-off frequency. By decreasing the cut-off frequency and masking more basal ANFs, the peak latency of the CAP response (N1) is delayed in a similar fashion to the cochlear traveling wave.

Generation of the $\Delta CAP(t)$ estimates. We recall the main equation (Eq 2): $\Delta CAP(t) = R * u(t)$, where $\Delta CAP(t)$ is a masking-release CAP waveform obtained by subtracting the response obtained under a given masking condition with the one obtained in the reference condition (no-notch). The right side of the equation is a convolution between a masking-release pattern $R(\tau)$ and a unitary response u.

Figure 2 describes the steps leading to the generation of the estimates $\widehat{\Delta CAP}(t)$. In the following, we justify these steps going backward from $\widehat{\Delta CAP}(t)$. $\widehat{\Delta CAP}(t)$ is obtained by convolution of a masking-release pattern $R(\tau)$ and the unitary response u. We assume that the masking-release pattern

R is related to the amount of masking M by

$$R = R_0(1 - M), \tag{4}$$

where R_0 is the difference in the excitation pattern between the full-notch condition (M=0; no forward-masker) relative to the no-notch condition (M=1; broadband noise). Here, we used the convention than $M = 1$ for the no-notch condition. As such, R_0 represents the maximum or fully unmasked masking release; its significance and how it is estimated is expressed later in the text. To generate the masking-release patterns, R_0 and M were first estimated in the CF domain by discretizing the linear frequency range [600 Hz, 12 kHz] into uniform intervals. This allowed us to define the masking-release pattern over CF, which was then converted into the latency domain by mapping latency and frequency bins using the power-law relationship mentioned in the previous paragraph. Finally, to compute the amount of masking M as a function of frequency, we relied on a simplified model of cochlear filtering using a linear filter bank. Given the average power spectral density of the masker $\mathcal{S}(f)$, the average response intensity I at the output of a cochlear filter was computed using:

$$< A^2 >= \int |w(f - CF)|^2 \, \mathcal{S}(f) \, \mathrm{d}f \,,$$

$$I = 10 \log_{10}(< A^2 >).$$

In the above equations, $< A^2 >$ is the intensity in linear units while I is in dB, and w is the cochlear filter defined in the frequency domain, centered around 0 and normalized such as its root mean square (RMS) value is 1. We considered that w was a 4th-order gammatone filter. The shape of w then depends only on the tuning of the cochlear filter at CF, characterized by the quality factor Q_{10} (related to the 10 dB-bandwidth by: $Q_{10} = CF/BW_{10}$). As the masker spectra are simple and defined by rectangular bands, analytical formulas for I as a function of CF were employed instead of integral expressions. These formulas are written in Appendix B.

After the computation of the response average intensities was done, the amount of masking M was finally obtained by applying the masking I/O function (Eq 3) to I.

2.3 Experimental Protocol

Preparation and anesthesia. To assess our method, we collected forward-masked CAP responses in 5 adult male chinchillas (*Chinchilla lanigera*) using surgical procedures pre-approved by the Purdue Animal Care and Use Committee. Anesthesia was induced using subcutaneous injections of xylazine (2-3 mg/kg) and ketamine (30-40 mg/kg). Anesthesia was maintained using intraperitoneal boluses of sodium pentobarbital (15 mg/kg/2h), and fluids (Lactated Ringer's) were administered subcutaneously throughout the experiment (∼1cc/hr). The animals' vital signs were monitored using pulse oximetry

(Nonin 8600V, Plymouth, MN) while oxygen was continuously delivered to the animal's nose area. Body temperature was maintained at 37°C using a homeothermic monitoring system with rectal probe (50-7220F, Harvard Apparatus).

Surgical procedure. Following anesthetic induction, a tracheotomy was performed to provide a low-resistance airway, reducing respiratory artifacts. Skin and muscles were transected following a dorsal-midline incision, and the external ear canals and bullae were subsequently exposed. Hollow ear bars were bilaterally placed in the ear canals and secured to a stereotaxic frame (David Kopf Instruments, Tujunga, CA). Sound was delivered monaurally through the ear bars using a dynamic loudspeaker (DT48, Beyerdynamic) at a sampling frequency of 48 kHz. To prevent a progressive negative pressure buildup in the bulla, a polyethylene tube (PE-90) was placed through an incision in the anterior bulla [23]. A second incision was made in the posterior base of the ipsilateral bulla to expose the middle ear. A silver wire electrode was placed near the round window to record CAPs and sealed in place within bulla opening using light-cured dental cement (Prime-Dent, USA). A pocket in the nape of the neck was made for a silver coiled wire reference electrode soaked in isotonic saline and connected to ground. All procedures were carried out in a double-walled, electrically shielded, sound-attenuating booth (Acoustic Systems, Austin, TX, USA). At the end of the experiments, animals were euthanized by barbiturate overdose.

Signal acquisition and pre-processing. We calibrated sound input using a probe microphone (Etymotic ER-7C) placed near the eardrum. A flat frequency response (within $\pm$ 2 dB until 10kHz) was achieved using a real-time 256-tap digital finite impulse response filter for the forward-maskers implemented using Tucker-Davis Technologies (TDT, Alachua, FL) hardware (RP2.1). For the click probe, we adopted a different equalization strategy by using the inverse of a 128th-order all-pole filter computed using linear predictive coding (LPC) to also correct for the phase differences induced by the acoustic system. CAP responses from the round window were amplified and band-passed using an ISO-80 Bio-Amplifier (10^3 gain, bandpass filtered from 10^2 to 10^4 Hz, World Precision Instruments) before being recorded by hardware modules (TDT RP2.1). Signal acquisition was controlled by a custom MATLAB-based (MathWorks, Natick, MA) interface. We used 5 chinchillas for this study, 4 of which had exploitable data at all center frequencies (CFs) tested (except at CF=8 kHz for chinchilla Q333). The last animal had exploitable data only within a limited frequency range (3–5 kHz), and is not included in the Results section, although the analysis we conducted on the partial data did not contradict the conclusions presented in the paper. Prior to analysis, the CAP responses were pre-processed by applying a Tukey window to isolate the time window where masking had a visible effect on the CAP (e.g., for chinchilla Q395: window defined on the interval [0.7, 5] ms, proportion of interval covered by the tapered cosine region: 0.4). The signals were smoothed by a Gaussian filter of standard deviation 0.03 ms (or, in frequency: 5.3 kHz). Additional

pre-processing was required in two animals to address specific experimental artifacts: for chinchilla Q395, a band-rejection filter was applied to remove a 1.5-kHz periodic electronic noise ; for chinchilla Q393, the DC components of the CAP responses were corrected to compensate for a slow DC drift.

Presentation of masker and probe. The relevant durations within one stimulus cycle are given in Figure 1a showing the time representation of the masker and probe. A cycle has a total duration of 160 ms. The durations were set according to existing data in the literature [24] as well as data collected during pilot experiments. We used in total 155 masking conditions, each associated with a one of the three power spectral profile shown in Fig 1b. Each stimulus cycle was repeated 120 times (12 blocks×10 repetitions). Within each block, the conditions were presented in a random order, ensuring some degree of interleaving to mitigate potential adverse effects due to long-term adaptation.

2.4 Masker design

A total of 155 masking conditions were presented during each experiment. Apart from the reference condition (broadband noise), the remaining 154 conditions were divided as follows: (a) high-pass noise maskers: n=12; (b) notched-noise maskers with a varying notch amplitude: n=77; (c) notched-noise maskers with a varying notch width, n=65.

The high-pass noise maskers (n=12) were each associated with a cut-off frequency ranging from from 1.2 kHz to 10 kHz. As mentioned before, these maskers were used to estimate the relationship between latencies and CFs.

The notched-noise maskers with a varying notch amplitude (n=77) were grouped according to the CF of the notch around 7 reference frequencies: 1.5, 2.2, 3, 4, 5, 6, and 8 kHz. For each CF, 10 maskers of this type were associated with a notch attenuation ranging from 35 dB to 0 dB, thus gradually merging into the no-notch condition (0-dB attenuation, reference condition). Except for the first experiment that was conducted (chinchilla Q395, corresponding to the data presented in the Results section), an additional condition was included with a notch of -3 dB attenuation (i.e., the power spectrum in the region of the notch was above the broadband-noise spectrum density); the introduction of this extra masker helped to determine the slope of the input-output masking curve at the reference power-spectrum level. The notched-noise maskers described in this paragraph typically had a large notch width (e.g., 2-kHz width at CF=5 kHz, 1-kHz width at CF=1.5 kHz). They were designed to estimate the amount of masking as a function of place-specific response intensity (masking I/O curves).

The remaining of the maskers were related to the third and last type: notched-noise maskers with a varying notch width (n=65). As for the previous type, these maskers were grouped according to the 7 reference frequencies. As an example, 10 maskers were associated to CF=5 kHz, with the notch width ranging from 900 Hz to 1.4 kHz, which is of the order of the expected value of the 10-dB bandwidth of cochlear filters at this CF [25]. To probe different groups of ANFs, the center frequency of the notch was intentionally put at

slightly different values between each masker; e.g., 4,800 Hz for one masker and 5,200 Hz for another. The notch amplitude for this type of masker was in most cases zero (infinite attenuation in dB). These maskers were designed to estimate the frequency selectivity of the cochlear filters. The approach is analogous to the measurement of critical bands in psychological studies, which also employ notched noise stimuli [2, 12, 13]. The underlying principle is based on the observation that, if the cochlear filters are sharply tuned, there is a rapid reduction in masking when the notch width is increased starting from the no-notch condition.

The frequency spectra of all the maskers were restricted to the range between 200 Hz and 12 kHz. The maximum power spectral density (PSD), corresponding to the sidebands of the maskers, was constant within each experimental session but varied across animals, ranging from 4 to 14 dB SPL. This range of maximum PSD corresponds to a sound level of 45 to 55 dB SPL for the no-notch (broadband noise) condition.

2.5 Estimation of model parameters

Model unknowns. The model introduced in the beginning of the Methods section and outlined in Fig 2 has multiple unknowns which are reviewed here:

1. The relationship between latencies and CFs. It is assumed to follow a power-law:
$$CF(\tau) = B(\tau - t_0)^{\alpha} \text{ for } \tau \geq t_0\,. \tag{5}$$
2. The unitary response u.
3. The amount of masking as a function of place-specific intensity response (*masking I/O curves*). In the case of the Weibull CDF (Eq 3), as adopted in the rest of the paper, this curve is parametrically defined by three variables (λ, I_0, s).
4. The tuning of the auditory filters, characterized by Q_{10}.
5. The distribution $R_0(f)$, which was introduced in Equation 4 as the fully unmasked masking release. As we move to more practical considerations, it is convenient to think of $R_0(f)$ as frequency weights that had to be included to account for the non-homogeneous contributions of different CFs to the masking release of the CAP.

Parameter estimation. This section presents the outline of the estimation of the parameters listed above. Technical details on the step-by-step procedure can be found in the code released for this project (jupyter notebook) [33]. To describe how the model was adjusted to the data, the model parameters can be separated into two groups. The first group of parameters are those that required a specific estimation procedure (described in the following paragraph): these are the parameters defining the place-latency relationship and the unitary response. The other parameters constituting the second group (masking I/O function, quality factor Q_{10}, and frequency weights) were estimated by minimizing the reconstruction error of the masking-release waveforms using a gradient descent algorithm.

The relationship between latencies and CFs was determined using the high-pass noise maskers; when presented in the order of decreasing cut-off frequencies, these maskers progressively mask the CAP from basal to more apical AN contributions. We considered that the ΔCAP peak delay (N1) was the latency associated with the cut-off frequency. The latencies were fit by a power-law estimated from the peak delays by least-squares fitting (Powell's dog leg method).

The unitary response u was estimated by deconvolution of the masking-release responses $[\Delta CAP(t)]$ with a first estimation of the masking-release patterns for the notched-noise maskers. For this step, the CAP responses were smoothed by a Gaussian filter of deviation 0.06 ms (or, in frequency: 2.6 kHz) instead of 0.03 ms elsewhere. Once the unitary responses and latencies were determined, they were considered fixed. However, after the estimation of all the parameters was done according to the procedures described in this sub-section, the unitary response was re-estimated with the updated masking-release patterns, and the estimation of the other parameters was performed a second time with the new unitary response.

Apart from the unitary response and the place-latency relationship, all the model parameters (highlighted in red in Fig 2) were fitted simultaneously by minimizing the mean squared error (MSE) between the signals $\widehat{\Delta CAP}(t)$ generated by the model, and the true signals $\Delta CAP(t)$. The implementation of the corresponding optimization algorithm is described in the paragraph *Optimization procedure* below.

One challenge of the method is that most of the model parameters potentially depend on CF. This is the case for the unitary response (depending on the normalized PSTH as defined in the model), the parameters controlling the masking I/O curve, the quality factor Q_{10}, and the frequency weights $R_0(f)$. This issue is mostly resolved by adjusting different versions of the model to each CF probed instead of having a single model fitted on all the data. For this purpose, the notched-noise maskers were grouped into 7 different center frequencies according to the frequency region of the notch (CF = 1.5, 2.2, 3, 4, 5, 6 or 8 kHz) and the associated forward-masked responses were fitted separately. However, the parameters defining the frequency weights $R_0(f)$ were shared across the different optimization processes. The estimation of $R_0(f)$ at every frequency was made possible at the cost of a regularity assumption considering that R_0 is a smooth function of f. We assumed that R_0 belongs to a low-dimensional manifold, explicitly that $R_0(f)$ in the range [200 Hz, 12 kHz] is only defined by its $m = 10$ first Fourier coefficients. For the estimation of Q_{10}, we assumed that the 10-dB bandwidth was constant in the interval of frequencies around CF and searched its optimal value using gradient descent. As an alternative, we also used a regression method assuming that Q_{10} could be approximated by a radial basis function (RBF) neural network. The input of the neural network was normalized frequency ($x = f/15000$) and its target was $\log Q_{10}$. The activations for the first layer were Gaussian functions with

standard deviation $\sigma = 0.5$). The first layer had 6 hidden neurons and the second layer (output) was a linear combination of the hidden neuron activations. If enabled, the RBF network was trained the same way as the other modules of the model, using gradient descent to minimize the reconstruction error of the masking-release waveforms.

Optimization procedure. The goal of the optimization procedure is to adjust the model parameters highlighted in red in Fig 2, including Q_{10} characterizing frequency selectivity, to obtain the best fit between the signals generated by the model and the true responses. While fitting all these parameters at once could seem intractable in a traditional setting, this approach is made possible by the fact that responses to many masking conditions are acquired during an experiment. We denote $[\Delta CAP(t)]_i$ the masking-release waveforms of the CAP, where i is an index for the masking condition ($i = 1 \cdots N_{\text{cond}}$, with N_{cond} being the total number of masking conditions). The model yields estimates $[\widehat{\Delta CAP}(t)]_i$ for each masking condition, and we define the cost function as the total mean square error:

$$MSE = \|\widehat{\Delta CAP} - \Delta CAP\|_2^2 = \sum_{i=1}^{N_{\text{cond}}} \sum_t \left([\widehat{\Delta CAP}(t)]_i - [\Delta CAP(t)]_i \right)^2$$

MSE was minimized by gradient descent. The gradients with respect to the model parameters were computed with PyTorch, an automatic differentiation library originally designed for the optimization of artificial neural networks [34]. A schematic for the graph of computations is provided in supplementary materials (Online Resource 2), that also synthesizes the operations that lead to the generation of $\widehat{\Delta CAP}(t)$. The key point is that, although the entire model is complex, each step of computation is a simple differentiable operation, and the gradients can be computed by applying the chain rule. An alternating gradient scheme was adopted. At step 1, the gradients were computed and summed over all the notched-noise masker conditions and the frequency weights $R_0(f)$ were updated. At step 2, the gradients were computed over the maskers with a notch of varying amplitude and the masking I/O function was updated. At step 3, Q_{10} was updated using the the maskers with a varying notch width. The same steps were then repeated about 100 times. The optimization was done separately for each CF probed. However, some parameters could be shared and optimized jointly – in particular, the frequency weights $R_0(f)$ – using the distributed communication package of PyTorch. The parameters were initially set manually or set at plausible values; e.g., Q_{10} was set to fit the curve $Q_{10} = 2(f/1000)^{0.5}$ loosely matching AN data [25], before being fine-tuned by the optimization algorithm. Since the cost function is not guaranteed to be convex with respect to the model parameters, and the algorithm can be stuck in local minima, several initializations were tried. For each run, a visual verification of the result was done; if the model did not achieve a close fit of the masking-release waveforms on the varying notch amplitude conditions, the initialization parameters were adjusted accordingly. Further discrimination

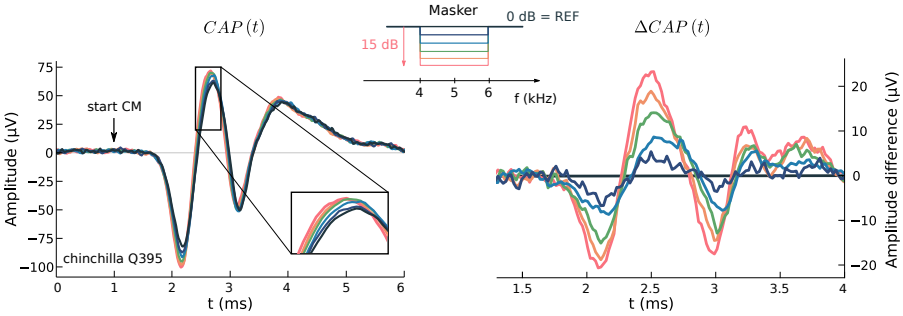

Fig. 3 Example of CAP data and derivation of $\Delta CAP(t)$. **Left:** forward-masked CAP responses to 80-dB (peSPL) clicks. The masker presents a 2-kHz notch of varying amplitude around 5 kHz (masker profiles are shown at the top center of the panel). **Right:** Corresponding masking releases, considering the no-notch condition as reference ($\Delta CAP(t) = CAP_{masked}(t) - CAP_{masked,b}(t)$). CM=cochlear microphonics. The notch attenuations goes from 15 to 0 dB (REF) in 3-dB steps.

between fine-tuned models was done by selecting the one associated with the minimum cost function.

3 Results

3.1 Estimation of input-output masking curves

Figure 3 shows an example of CAP responses in the presence of forward-maskers with a notch of varying amplitude. The right panel of the figure shows the corresponding $\Delta CAP(t)$ waveforms derived from the forward-masked CAPs by subtracting the reference response. The maximum peak-to-peak (p-p) amplitude of the masking-release ΔCAP is approximately a third of the p-p amplitude of the baseline masked CAP ($CAP_{masked,b}$; response associated with the no-notch condition), which in turn accounts for about half of the p-p amplitude of the unmasked CAP (recorded in the absence of a forward-masker; not shown). A first indication of the masking input-output curve – the amount of masking as a function of cochlear-filter output intensity – is provided by the measure of reduction of the ΔCAP p-p amplitude when the masker notch attenuation is progressively decreased (Fig 4a). In reality, the relationship between the reduction of the CAP peak amplitude and the underlying masking I/O curve is not guaranteed to be linear, because the masking of the CAP also depends on the spread of the cochlear excitation pattern, which differs for each masker. For this reason, the determination of the reduction of the CAP amplitude serves only as a first approximation of the parametric masking I/O curve, which is then fine-tuned during the optimization procedure along with the other parameters. The masking I/O curve at CF=5 kHz after optimization (dashed line) is also shown in Fig 4a, clearly deviating from the initial curve. The other curves for the same animal at different CFs are shown in panel B. We did not find a regular pattern in the changes of the I/O

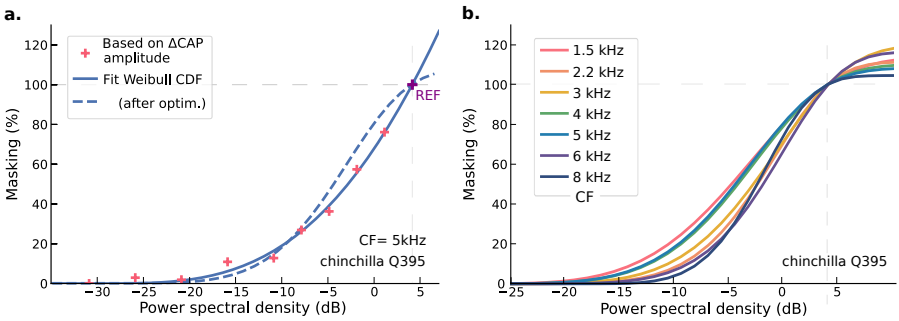

Fig. 4 Masking input-output (I/O) curves. **a.** Amount of masking at CF=5 kHz as estimated by the peak-to-peak (p-p) amplitude of the responses represented in Fig 3 (masker with a 2 kHz-wide notch centered at 5 kHz). The x-axis refers to the power spectral density in dB SPL within the notch. The purple cross corresponds to the reference condition (i.e., no-notch condition, matched to 100% masking). The fit to this data using a Weibull CDF is shown, as well as the fit after fine-tuning the model (dashed line; see text for discussion of why this curve is different from the p-p amplitude data), considered to better approximate the underlying masking I/O function of the compound response of ANFs tuned to CF. **b.** Masking I/O curves (Weibull CDFs) for the same animal at every CF after fine-tuning the model.

curves with CF considering all the animals in the study. Note that since the I/O functions were computed using the notched-noise maskers, the amount of 0% masking does not necessarily mean that no masking occurred for that level but rather that there was no additional masking relative to the minimum masking condition (masker presenting a wide notch with 35-dB attenuation).

3.2 Estimation of latencies and frequency weights

Figure 5 shows the estimated latencies for the same chinchilla using the narrowband analysis method. Although the relative errors appear to be larger at high frequencies, it is the deviations from the power law at lower frequencies (deviations of 0.15 ms for this animal at 1 kHz, up to 0.3 ms in another animal) which have a greater impact on the model performance. Note that to obtain the CAP peak latencies, it is necessary to also take into account the peak delays of the estimated unitary response, shown for the same animal in Fig 6a. The estimate unitary response u keeps the biphasic shape of the spike unit response typically reported [31], but is repeated at least twice, with the two first negative peaks separated by 0.8 ms. The second peak has been seen in other studies and partly attributed to the phenomenon of 'double-spiking', i.e., the firing of ANFs immediately after the refractory period [28, 35]. However, this phenonemon is not systematically seen in the PSTH of ANFs in response to clicks [36, 37]. Another reason may be the presence of sub-threshold electrical resonances in the auditory nerve peripheral dendrites [38]. Interestingly, this figure does not exhibit significant variations in the shape of the unitary response, but small changes with a trend consistent with decreasing CFs can be observed at 2.2 and 3 ms. These changes could be explained by larger group delays for apical cochlear filters (i.e., a slower build-up of response intensity),

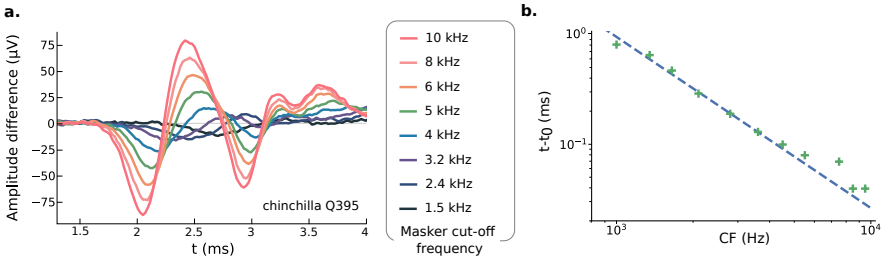

Fig. 5 Estimation of the place-latency relationship. **a.** Masking-release waveforms ΔCAP for the high-pass noise maskers. The cut-off frequency goes from 10 kHz to 1.5 kHz (REF: broadband noise, cut-off frequency 200 Hz). The responses display a shift in the peak latency (N1) that follows the same trend as the cochlear traveling wave. The responses for 8 out of 12 high-pass noise maskers are shown. **b.** Estimated latencies as a function of CF (green crosses, log-log scale) from the data represented in panel A. Fit (dashed line): power law, $CF = 11.6\,(t - t_0)^{-0.64}$, with $t_0 = 0.83$ ms (standard error: 0.05 ms; note: to interpret the value of t_0, one also has to take into account the timing of the unitary response, see Fig 6a).

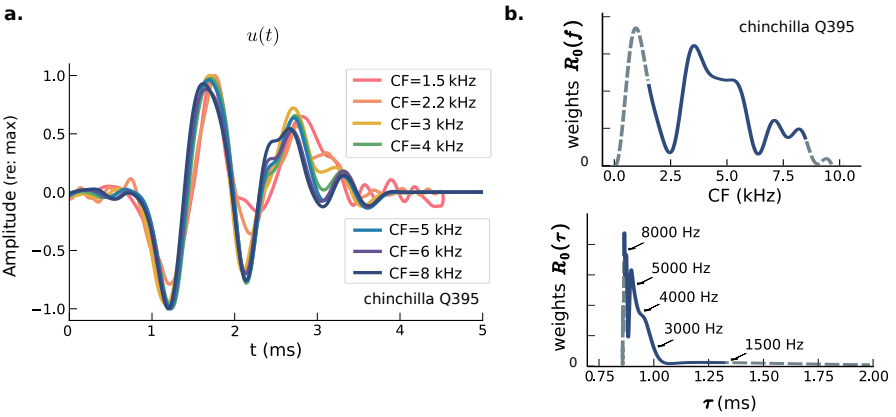

Fig. 6 Other ancillary parameters of the model. **a.** Estimated unitary responses u corresponding to the weighted average of deconvolutions of ΔCAP responses (notched-noise maskers with a varying notch attenuation) with their associated masking-release patterns. The estimated unitary responses have been normalized to have the same baseline-to-peak amplitude. **b.** Estimation of the frequency weight distribution $R_0(f)$ (top) representing the relative CF contributions to ΔCAP. The weights below 1.5 kHz and above 8kHz (dashed lines) are a result of extrapolation and do not correspond to real data points. The associated distribution in the latency domain is shown (bottom). The conversion from CF to latency was done using the relation $CF(\tau) = B(\tau - t_0)_+^{\alpha}$, with the change of variable $R_0(f)df = R_0(f)B\alpha(\tau - t_0)^{\alpha-1}d\tau = R_0(\tau)d\tau$.

hence a prolonged peak in $n\Delta PST$ for lower CFs in Equation A2. A fast analysis based on the deconvolution of the unitary responses at each CF with the unitary response at CF=8 kHz supports this hypothesis.

Fig 6b shows the estimation of the frequency weights $R_0(f)$ representing how different CFs contribute to ΔCAP relative to each other. The distribution of weights is also shown as a function of latencies using a change of variable

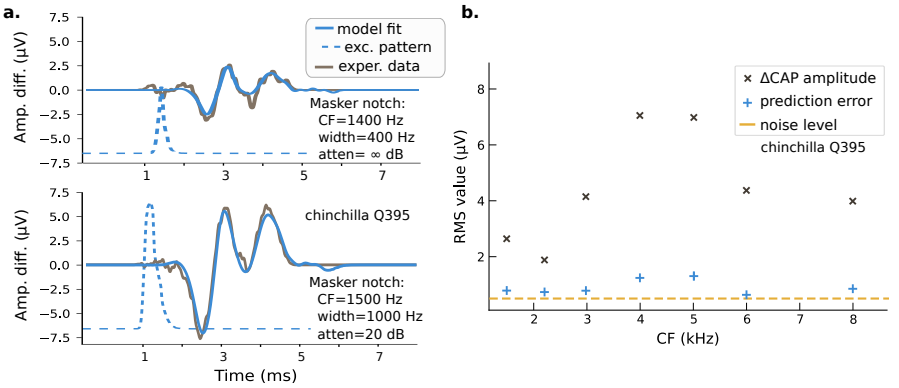

Fig. 7 Fitting of $\Delta CAP(t)$: results. **a.** Two examples of fits of $\Delta CAP(t)$ for two notched-noise maskers after parameter optimization. The first masker belongs to the varying notch width type, while the second masker belongs to the varying notch attenuation type. Masking-release excitation patterns are shown in dashed blue (arbitrary scale and zero for y-axis). **b.** Synthesis of errors and ΔCAP RMS amplitude values (computed on the 100% region of the Tukey window after pre-processing of the data) at the different CFs from the same animal. The squared errors are averaged across all conditions corresponding to notched-noise maskers with a notch centered around CF.

(panel B, bottom). The gradual decrease of $R_0(f)$ with CF was expected since the distribution of the preferred frequency of ANFs is denser at low frequencies as a result of the exponential relationship between cochlear place and CF. However, as shown in Fig 6b, $R_0(f)$ exhibits in addition two narrow dips (2.5 kHz and 6 kHz) that hinder the estimation, not only of the frequency weights, but also of the other parameters of the model at the corresponding frequencies. In the same time, since $R_0(f)$ is estimated as a sum of sine and cosine functions, oscillations in the approximation of $R_0(f)$ can affect the prediction of other model parameters. To deal with this issue, we adopted a strategy consisting in approximating $R_0(f)$ with low Fourier modes only ($m = 4$) at the initialization of the optimization procedure, then increasing the maximum mode ($m = 10$) during gradient descent. Most of the chinchillas presented the same type of distributions, with an overall decreasing trend for $R_0(f)$ and one or two relatively narrow dips, but the dips were not always found at the same frequencies across animals. We do not have a definitive explanation for the presence of dips in $R_0(f)$, but a speculative hypothesis is that they result from the three-dimensional spiral cochlear geometry. Previous studies have suggested that the geometrical configuration of the cochlea could account for the presence of dips in the spatial contributions to the cochlear microphonic [39]. However, the lack of understanding regarding the spatial origin of the CAP adds an additional layer of uncertainty to this explanation.

3.3 Fitting of ΔCAP and estimation of frequency selectivity

Figure 7a shows how the model fit the experimental data for two masking conditions. Panel B is a synthesis of the prediction errors as a function of CF for the same animal. In most cases, more than 90% of the variance was accounted for by the model. Remarkably, for some CFs, the prediction error almost reached noise level (after pre-processing). Equally robust fits were also obtained for the other animals in the study.

Finally, we present the results of the estimation of frequency selectivity, which was the main goal of the present study. Fig 8 shows the fitting error for ΔCAP for CF=6 kHz as a function of model filter 10-dB bandwidth for all chinchillas. The fitting error is computed on the the maskers presenting a varying notch width around CF=6 kHz which were designed to estimate cochlear frequency selectivity. To plot this figure, the model parameters were found by gradient descent and considered fixed except for the 10-dB bandwidth which was varied from 500 Hz to 5 kHz. The bandwidth minimizing the prediction error provides an estimate of the 10-dB bandwidth at CF, as shown by the arrow for one of the animals. Two curves exhibit a larger curvature at their minimum point, showing that the data collected from different animals do not always provide the same amount of information about Q_{10} (in the sense of Fisher information). For each chinchilla, we also estimated Q_{10} as a function of CF using a RBF network, to take advantage of the assumed smoothness and regularity of the quality factor with respect to CF. These regressions are shown in Fig 8b, along with their average and standard deviation across animals. An average of Q_{10} values derived directly from AN tuning curves is also provided for comparison [25], highlighting the close match between the two datasets.

4 Discussion

4.1 Suitability of the convolution model for forward-masked CAPs

Our approach to fit forward-masked CAP responses with a differentiable convolution-based model led to accurate predictions of forward-masked CAP waveforms. The generation of the waveform estimates relied on a consistent set of parameters (including cochlear frequency selectivity), which were estimated by gradient descent (parametric masking I/O function, frequency weights, quality factor Q_{10}) or by a specific procedure (latencies and unitary responses). More than 90% of the variance of the masking-release waveform $\Delta CAP(t)$ was explained by the model considering the Gaussian notched-noise maskers, in most animals and CFs. Note that this prediction error is calculated without cross-validation, so that a part of the performance of the model could be due to overfitting the data. However, the number of maskers divided by the number of reference CFs (155/7=22.14) is relatively large compared to the number of effective parameters by CF ($\sim$ 6, excluding the unitary response estimated

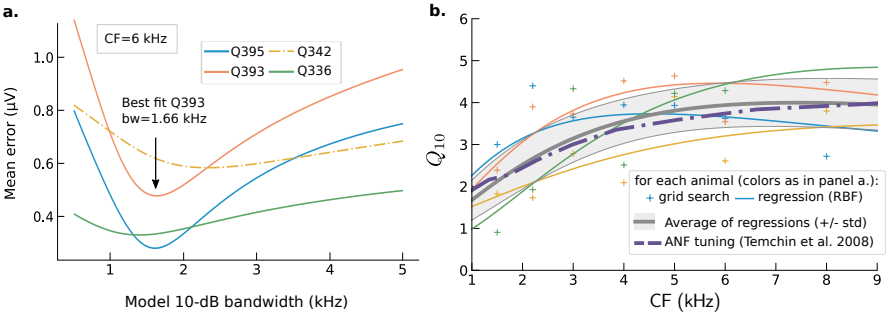

Fig. 8 Estimation of the frequency selectivity. **a.** Plot of the fitting errors of $\widehat{\Delta CAP}$ as a function of model 10-dB bandwidth for each chinchilla. The minimum error corresponds to the estimate of the 10-dB bandwidth using gradient descent (e.g., the estimated bandwidth for Q393 at 6 kHz is 1.66 kHz). The error was computed over the responses corresponding to the notched noise maskers presenting a varying notch width around CF=6 kHz. The data for chinchilla Q432 (dashed line), characterized by a lower signal amplitude, was multiplied by 5 to match the scale of the other plots. **b.** Synthesis of the estimates of the quality factor Q_{10} as a function of CF. Crosses correspond to the estimates using gradient descent for each CF independently; solid lines correspond to estimates using a regression technique (RBF network) during the optimization of the model parameters. The gray shaded area shows the average and standard deviation of the regressions (solid lines). Average data from published ANF recordings in chinchillas [25] are given for comparison (dashed purple line) to support the accuracy of the current approach.

from a weighted average of the CAP waveforms). The interactions between the variables of the model are also limited thanks to the design of each type of masker to estimate specific parameters (i.e, the high-pass noise maskers are used to estimate the parameters of the place-latency model; the maskers with a varying notch amplitude are used to estimate the masking I/O function, etc.). These two factors combined reduce the potential for overfitting.

One of the underlying assumptions of the convolution model was that applying different degrees of masking by manipulating the masker notch would substantially change the amplitude of the masking-release waveform ΔCAP, but not its overall shape (Appendix A). We found that it was indeed the case, as shown for example in Fig 3 (right). However, we noticed an exception during a pilot experiment where a reduction in the masker notch attenuation from 15 dB to 6 dB resulted in an additional 0.1 ms delay in ΔCAP. We attributed this observation to the lower probe sound level that was used for this experiment compared to subsequent sessions, suggesting the importance of using a loud probe to obtain a sufficiently time-localized PSTH and mitigate this issue.

In the convolution model, the latencies were related to CF by a power-law. Although it captured the overall trend well, the local dependence of latencies on CF was not always properly described by a single power-law model fitted over the entire range of CFs. In addition, the latencies were small above CF=4 kHz (<0.1 ms, Fig 5b) relative to the width of the unitary response, therefore the relevance of the convolution model could be questioned for high CFs. A simpler model in which all the contributions of higher CFs are considered

synchronous would probably equally well describe the ΔCAP masking-release waveforms. However, the convolution model remains robust when the latencies are of small magnitude and can provide a more accurate model for lower CFs where the latency differences are more pronounced. Latencies could also show greater variations when considering other animal species.

The fitting of the ΔCAP waveforms after the adjustment of the model was remarkably accurate, but the estimation procedure presented several challenges. The fact that the model relies on a relatively large number of parameters, especially if we include all the possible dependencies on CF, can make the optimization cumbersome. The optimization of the model is however facilitated by the existence of new elegant libraries for automatic differentiation. We found that the main difficulty regarding the estimation of the different model parameters was the determination of the frequency weights $R_0(f)$. The model would be greatly simplified if we could assume that different CFs contribute to ΔCAP with the same magnitude, but we found that it was not the case. We showed one extreme case in Fig 6b, where two narrow dips (at 2.5 kHz and 6 kHz) were present in $R_0(f)$. The estimation of $R_0(f)$ is still possible with regularity assumptions and notched-noise maskers with notches distributed over the entire range of frequencies. However, if the dips are too steep, the estimation of the frequency weights and of the other parameters can be affected. As potential evidence, the largest deviation between Q_{10} values derived from AN tuning curves and those obtained with our estimation procedure (Fig 8b) was observed at 2.2 kHz for the animal presenting a dip around this frequency (blue cross). By using a regression technique for the estimation of Q_{10}, we can however exploit the regularity of the quality factor with respect to CF to still provide a reasonable estimate of frequency selectivity (blue solid line in Fig 8b).

4.2 Estimation of frequency selectivity using forward-masked CAPs

We found a good agreement between the estimates of the quality factor averaged over the 4 experiments for which we had complete data (Fig 8) and published values derived from the collection of many ANF tuning curves in chinchillas [25]. For one animal (chinchilla Q395), single-fiber auditory-nerve recordings were conducted for another experiment after the end of the collection of the CAP responses. The tuning-curve Q10 factors from the ANF recordings were close to the Temchin et al. (2008) data, but the estimate of Q10 using forward-masked CAPs followed a non-monotonic trend not seen in the data (blue line, Fig 8b). This suggests that, while the results matched published ANF data when averaged over experiments, the method is not robust enough to provide an exact estimate at an individual level. Overall, it is necessary to exercise some caution when interpreting the results, despite the striking similarity observed between our averaged estimates and the averaged data from Temchin et al. (2008). Indeed, several factors add some uncertainty to the comparison of quality factor estimates: a) there is a significant spread

of Q_{10} values when examining individual ANF tuning curves [40], and their measurements typically involve the use of low-intensity tones; b) tuning-curve estimates employ an iso-response method, while our approach is closer to an iso-input method: this is known to have consequences for the estimation of frequency tuning [6, 41]; c) the number of experiments conducted here was limited to 4. Given these factors, the main conclusion is that the two averaged sets of data fall in the same range of values and have the same CF trends, rather than indicating outstanding accuracy in predicting absolute Q_{10} values using our method. Despite these potential limitations, the accuracy of our method in overall values suggests that it may be highly valuable as an evoked-response method for comparing AN tuning across hearing-loss etiologies and/or across species without requiring single-unit data.

Our experimental approach is akin to the experiments of Verschooten and colleagues [4, 18] on the estimation of cochlear tuning – their work was in turn an improvement of experiments using forward-masked CAPs that were conducted in the 1980s [15, 17]. The estimation method used by Verschooten et al. involved establishing iso-response curves for masker level versus masker notch width – the response criterion being that 66% of the initial CAP amplitude had to be restored. A measure of tuning was derived from these curves by considering the 10-dB bandwidth – reduced to a single auditory filter model, this measure can be seen as the bandwidth encompassing 90% of the frequency response power spectrum (called BW90 in other works [42]). The main advantage of their technique compared to ours is that it did not require the assumption that the amount of masking of synchronized ANFs is driven by input-output curves that are to be determined. Rather, their measure of tuning was considered as an empirical quantity, and assumed to be proportional to the 10-dB bandwidth of ANF tuning curves. They found a good agreement between the two quantities after a constant correction factor was applied. However, the conversion factor from CAP to ANF data was not the same for every species and was smaller for small mammals. In addition, the correction factor for macaques was not constant as a function of CF (S5 Fig in [4]). It is therefore not clear how the derived measure can be interpreted, as it may be affected differently from one species to another by different factors (e.g., differences in masking-release patterns). By contrast, the strength of our method lies in the mathematical modeling of the forward-masked CAPs that seeks to capture the essential physiological aspects relevant to the amplitude and shape of the CAP waveforms. Our model incorporates the complex relationship between frequency selectivity and the effect of masker manipulations (e.g., increasing notch width) on the forward-masked CAP waveforms. This limits the reliance on an empirical correction factor. We were able to fit ANF tuning data without any additional factor (Fig 8) on the experimental data we collected. Further testing of our method's accuracy would require additional experimental work and analysis involving other species. The convolution-based method could have other advantages. Since the entire ΔCAP signal is used instead of only the CAP peaks, one can expect the method to be more robust

to noise. Furthermore, it exploits all the available data, whereas the 'fast' procedure in Verschooten et al. searches for a particular masker level meeting the masking criterion, thus potentially wasting measurement points. Beyond these aspects, a potential of our analytical approach is that the mathematical model and experimental paradigm could be adapted to study more complex aspects of cochlear signal processing, such as compressive nonlinearities, as mentioned in the next paragraph.

4.3 Limitations related to the simplified underlying auditory model

A few difficulties associated with the model were mentioned throughout the paper, including changes in the model parameters with CF that make estimation more challenging. Another set of limitations is related to the oversimplifications of the model to describe the behavior of the cochlea. One major drawback of the model is that it assumes that the cochlear frequency decomposition, implemented by a filter bank independent of sound level (Fig 2), is linear. However, this assumption is not valid for the healthy ear, since compressive nonlinearities decrease the cochlear frequency selectivity when intensity is increased [5]. The non-linearities also modify the input-output functions depending on the amount of suppression [14]. Therefore, including these nonlinear effects in the model or adapting more complete computational models of the auditory periphery for the proposed paradigm (e.g., BEZ model [29]) could provide insights into how compressive nonlinearities affect cochlear processing, but given the major extension in scope, is left for future developments. To study nonlinear effects, a greater variety of masking conditions would have to be employed during data collection (e.g., various level or asymmetry of notches relative to one CF); however, this additional set of conditions is certainly possible. As an example, Verschooten et al. evaluated the level dependence of cochlear frequency selectivity in cats by presenting maskers of various intensities [18].

Other aspects of the auditory model considered in this work also correspond to oversimplifications of cochlear signal processing. Auditory filter frequency profiles are in reality asymmetric, and the lower and upper sides are not affected the same way by nonlinearities [43]. We focused on the 'tip' of the auditory filters, which can be accurately described by gammatones – we also tried Gaussian filters and did not find significant differences using one model or the other. But auditory filters also present a low-frequency tail, the latter showing different attributes depending on filter CF [25]. In addition, the CF of cochlear filters change with the degree of compression [44]. Future work is needed to explore whether the proposed method could be extended to include these different aspects.

Compared to non-invasive techniques for measuring cochlear frequency selectivity (such as OAE-based methods or psychoacoustic experiments based on masking), estimation methods using forward-masked CAPs have the advantage of being more closely related to AN activity. This advantage would

be especially meaningful if more complex aspects of cochlear processing, such as compressive nonlinearities, could be integrated in the model. In the transition of the proposed method to human subjects, the question of the quality of the CAP measurements also gains importance. While recent studies using tympanic membrane show promise for obtaining a reasonable SNR [45, 46], extra-tympanic measurements alone may be too limited to provide exploitable data. More invasive alternatives also exist for translating the method to humans, such as trans-tympanic measurements [4, 11] or intracranial recordings during surgeries for neurovascular conflicts [47–49].

Supplementary information.

- **Online Resource 1**: Simulation of the effect of masking on the compound response of AN fibers. A/ Post-stimulus time histograms (PSTHs) of a population of ANFs using the Bruce et al. model (2018). The bar plots are the compound PSTHs in response to a 90 dB SPL click presented 5 ms after a 30-ms white Gaussian noise masker with a masker level varying from 10 dB to 80 dB. Simulation parameters: CF = 4 kHz, population of 32 ANFs (low spontaneous rate LS: 6, MS: 6, HS: 20); N reps = 400 (clicks were of alternating polarity); bin interval: 0.1 ms. B/ Difference of the compound PSTHs shown in a/ taking as reference the response to masker with the lowest level. C/ Difference in spike rate for four masker sound levels reproduced from the simulated PSTHs, this time with the highest masker intensity as reference. When normalized, the PSTH differences correspond to $n\Delta PST$ in Appendix A (this figure shows that the ΔPSTs have similar shapes but different amplitudes). D/ Convolution of the PSTHs differences (shown in C/) with the spike unit response in Wang 1979 (Fig 14). It simulates ΔCAP for different masker sound levels, however considering only one frequency channel (CF=4 kHz).
- **Online Resource 2**: Diagram of computations (divided in two parts A and B) for a masker with 2 bands implemented using PyTorch leading to the generation of the estimated CAP masking-release $\widehat{\Delta CAP}(t)$.The variables that are updated during the optimization procedure (gradient descent) are represented in gold font.

Acknowledgments. This work was conducted while the first author was on a postdoctoral fellowship supported by Fondation Pour l'Audition (FPA RD-2019-3). Support was also provided by NIH grants R01-DC009838, T32-DC016853 and F30-DC020916.

Declarations

Conflict of interest: The authors declare that they have no conflict of interest.

Appendix A Breakdown of the convolution equations

In this appendix, we take a closer look at the convolution equations (Equations 1 and 2), since the assumptions justifying these formulas were implicit in the main body of the paper.

We first start by considering the equation for the non-masked version of the CAP (Eq 1), recalled here:

$$CAP\,(t) = E * u_0\,(t)\,.$$

This formula can in fact correspond to two different approaches, depending on how the unitary response is defined. If the unitary response is the same as the spike unit response, then E is the compound post-stimulus time histogram (PSTH) of all ANFs. This is consistent with several studies that use elaborate computational models of ANF activity to simulate compound PSTHs then generate CAP waveforms [20, 50]. This differs however from our approach, as we actually never try to reproduce the compound PSTH. In this sense, our work is closer to a second approach [26, 27], where E is defined as an excitation spread over latencies with a one-to-one mapping of latencies and CFs. This approach necessarily requires that the spike-time jitter of ANFs tuned to a given CF is encompassed in the unitary response since it cannot be included in the excitation pattern.

To make the above distinction more explicit, let us consider $nPST$, the compound PSTH of ANFs tuned to a single CF normalized by the total number of spikes. We also assume that only a limited segment of the cochlear partition contributes to the CAP waveform, so that $nPST$ can be considered independent of CF. We can then write a *double* convolution equation for the CAP waveform:

$$CAP(t) = \underbrace{E * \overbrace{nPST}^{u_0} * UR}_{cPST}(t) \tag{A1}$$

where $nPST, cPST$ stand for the normalized and compound PSTHs, and UR is the spike unit response. In this formula, we have used notations similar to Bappert et al. [27], which describes the double convolution approach in more details. If $nPST$ is left outside the function to the right of the convolution, the unitary response aligns with the spike unit response, corresponding to the first approach of simulating the compound PSTH. In our case, however, the unitary response u_0 takes into account the spike unit response as well as the normalized spike histogram $nPST$.

Now let us take a closer look at the equation for the masked version of the CAP, specifically the masking-release $\Delta CAP(t)$, which is the real focus of this paper. Since we consider masking releases associated with the manipulation of a narrow spectral notch, the assumption that only a limited segment of the

cochlear partition contributes to $\Delta CAP(t)$ is always justified. We can therefore approximate $\Delta CAP(t)$ with an equation similar to Eq A1:

$$\Delta CAP(t) = \underbrace{R * \overbrace{n\Delta PST * UR}^{u}}_{c\Delta PST}(t) = R * u(t)\,. \tag{A2}$$

We recall that $R(\tau)$ is the masking-release pattern and u is the unitary response – here, the zero subscript has been removed to distinguish the unitary response from the one in Eq A1, defined differently. Again, u is considered as the compound of the spike unit response and the difference in the PSTH of a population of synchronized ANFs normalized with respect to the amount of masking ($n\Delta PST$). If, on the other hand, the unitary response was identical to the spike unit response UR, the function to the left of the convolution would be the difference in the compound PSTH induced by masking ($c\Delta PST$). We are not interested, however, in the actual decomposition of u, justifying that the simpler equation $\Delta CAP = R * u$ is kept in the main body of the paper. It is worth noting that, in this equation, we assume that $n\Delta PST$ is independent of the amount of masking. If a click probe of medium-to-high intensity is used, the individual PSTHs are characterized by a sharp predominant peak restricted on a short time interval [28]. As a result, the changes in the shape of $n\Delta PST$ are expected to have a minimal effect on the CAP; however, the amount of masking applied to the PSTHs will have a significant impact on the CAP amplitude. Prior to any experiment, we tested whether this hypothesis was reasonable with a well-established computational model of ANF responses (BEZ model [29]). This analysis is left as supplementary material (Online Resource 1). As for the spike unit response UR, studies have typically reported that it can be considered independent of the ANF best frequency or spontaneous rate [30–32].

Appendix B Gammatone model

This appendix contains the computation of the average intensity at the output of a gammatone cochlear filter:

Note: In this paragraph, τ does not have the same use as in the main part of the paper where it is a variable for latencies. Here, it refers to the time constant of the gammatones.

The k-th order gammatone, characterized by an envelope proportional to $t_+^{k-1}e^{-t/\tau}$, is defined in the frequency domain (complex version, w.l.o.g.) by:

$$|w_{CF}(\omega)|^2 = \binom{2k-2}{k-1}^{-1} 2^{2k-1}\tau\left[1+\tau^2(\omega-2\pi\,CF)^2\right]^{-k}\,.$$

The average quadratic response considering a single-band Gaussian-noise masker is:

$$< A^2 >= S_0\binom{2k-2}{k-1}^{-1} 2^{2k-2}\pi^{-1}\tau \int_{2\pi(f_{\min}-CF)}^{2\pi(f_{\max}-CF)} \left[1+\tau^2\omega^2\right]^{-k} d\omega$$

$$< A^2 >= S_0 \binom{2k-2}{k-1}^{-1} 2^{2k-2}\pi^{-1} \int_{\arctan(2\pi\tau(f_{\min}-CF))}^{\arctan(2\pi\tau(f_{\max}-CF))} \cos^{2(k-1)}\theta \, d\theta \ .$$

The last integral is then computed by writing
$\cos^{2(k-1)}\theta = 2^{2-2k}\left[\sum_{l=0}^{k-2}\binom{2k-2}{l} 2\cos((2k-2-2l)\theta) + \binom{2k-2}{k-1}\right]$.

In the case of a masker presenting multiple bands, the expressions for each band simply add up.

Note: The 10-dB bandwidth is related to τ *by* $BW_{10}\tau\pi = \left[10^{1/k}-1\right]^{1/2}$.

References

[1] Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proceedings of the National Academy of Sciences 99(5):3318–3323. https://doi.org/10.1073/pnas.032675099

[2] Oxenham AJ, Shera CA (2003) Estimates of human cochlear tuning at low levels using forward and simultaneous masking. Journal of the Association for Research in Otolaryngology : JARO 4(4):541–54. https://doi.org/10.1007/s10162-002-3058-y

[3] Sumner CJ, Wells TT, Bergevin C, et al (2018) Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. Proceedings of the National Academy of Sciences of the United States of America 115(44):11,322–11,326. https://doi.org/10.1073/pnas.1810766115

[4] Verschooten E, Desloovere C, Joris PX (2018) High-resolution frequency tuning but not temporal coding in the human cochlea. PLOS Biology 16(10):e2005,164. https://doi.org/10.1371/journal.pbio.2005164

[5] Heinz MG, Colburn HS, Carney LH (2002) Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates. The Journal of the Acoustical Society of America 111(2):996–1011. https://doi.org/10.1121/1.1436071

[6] Eustaquio-Martín A, Lopez-Poveda EA (2011) Isoresponse versus isoinput estimates of cochlear filter tuning. JARO - Journal of the Association for Research in Otolaryngology 12(3):281–299. https://doi.org/10.1007/s10162-010-0252-1

[7] Ruggero MA, Temchin AN (2005) Unexceptional sharpness of frequency tuning in the human cochlea. Proceedings of the National Academy of

Sciences of the United States of America 102(51):18,614–18,619. https://doi.org/10.1073/pnas.0509323102

[8] Shera CA, Charaziak KK (2019) Cochlear frequency tuning and otoacoustic emissions. Cold Spring Harbor Perspectives in Medicine 9(2). https://doi.org/10.1101/cshperspect.a033498

[9] Wilson US, Browning-Kamins J, Durante AS, et al (2021) Cochlear tuning estimates from level ratio functions of distortion product otoacoustic emissions. International Journal of Audiology https://doi.org/10.1080/14992027.2021.1886352

[10] Eggermont JJ (2017) Ups and downs in 75 years of electrocochleography. Frontiers in Systems Neuroscience 11:2. https://doi.org/10.3389/fnsys.2017.00002

[11] Verschooten E, Joris PX (2022) Measurement of Human Cochlear and Auditory Nerve Potentials. In: Groves AK (ed) Developmental, Physiological, and Functional Neurobiology of the Inner Ear. Humana, New York, NY, chap 14, p 321–337, https://doi.org/10.1007/978-1-0716-2022-9_14

[12] Patterson RD (1976) Auditory filter shapes derived with noise stimuli. The Journal of the Acoustical Society of America 59(3):640–654. https://doi.org/10.1121/1.380914

[13] Moore BC, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. Journal of the Acoustical Society of America 74(3):750–753. https://doi.org/10.1121/1.389861

[14] Delgutte B (1990) Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers. Journal of the Acoustical Society of America 87(2):791–809. https://doi.org/10.1121/1.398891

[15] Harrison RV, Aran J, Erre J (1981) AP tuning curves from normal and pathological human and guinea pig cochleas. The Journal of the Acoustical Society of America 69(5):1374–1385. https://doi.org/10.1121/1.385819

[16] Charaziak KK, Siegel JH (2014) Estimating Cochlear Frequency Selectivity with Stimulus-frequency Otoacoustic Emissions in Chinchillas. JARO - Journal of the Association for Research in Otolaryngology 15(6):883–896. https://doi.org/10.1007/s10162-014-0487-3

[17] Harrison RV, Aran JM, Negrevergne M (1981) The Frequency Selectivity of the Normal and Pathological Human Cochlea. Arch Otorhinolaryngol 230(3):221–227. https://doi.org/10.1007/BF00456322

[18] Verschooten E, Robles L, Kovačić D, et al (2012) Auditory nerve frequency tuning measured with forward-masked compound action potentials. JARO - Journal of the Association for Research in Otolaryngology 13(6):799–817. https://doi.org/10.1007/s10162-012-0346-z

[19] Goldstein MH, Kiang NY (1958) Synchrony of Neural Activity in Electric Responses Evoked by Transient Acoustic Stimuli. Journal of the Acoustical Society of America 30(2):107–114. https://doi.org/10.1121/1.1909497

[20] de Boer E (1975) Synthetic whole-nerve action potentials for the cat. Journal of the Acoustical Society of America 58(5):1030–1045. https://doi.org/10.1121/1.380762

[21] Eggermont JJ (1976) Analysis of compound action potential responses to tone bursts in the human and guinea pig cochlea. Journal of the Acoustical Society of America 60(5):1132–1139. https://doi.org/10.1121/1.381214

[22] Prijs VF, Eggermont JJ (1981) Narrow-band analysis of compound action potentials for several stimulus conditions in the guinea pig. Hearing Research 4(1):23–41. https://doi.org/10.1016/0378-5955(81)90034-4

[23] Guinan JJ, Peake WT (1967) Middle-Ear Characteristics of Anesthetized Cats. The Journal of the Acoustical Society of America 41(5):1237–1261. https://doi.org/10.1121/1.1910465

[24] Abbas PJ, Gorga MP (1981) AP responses in forward-masking paradigms and relationship to responses of auditory-nerve fibers. The Journal of the Acoustical Society of America 69(2):492–499. https://doi.org/10.1121/1.385477

[25] Temchin AN, Rich NC, Ruggero MA (2008) Threshold tuning curves of chinchilla auditory-nerve fibers. I. Dependence on characteristic frequency and relation to the magnitudes of cochlear vibrations. Journal of Neurophysiology 100(5):2889–2898. https://doi.org/10.1152/jn.90637.2008

[26] Elberling C, Hoke M (1978) Brief communication: Decoding of human compound action potentials. Scandinavian Audiology 7(3):171–175. https://doi.org/10.3109/01050397809076284

[27] Bappert E, Hoke M, Lütkenhöner B (1980) Deconvolution of compound action potentials and nonlinear features of the PST histogram. Hearing Research 2(3-4):573–579. https://doi.org/10.1016/0378-5955(80)90095-7

[28] Versnel H, Schoonhoven R, Prijs VF (1992) Single-fibre and whole-nerve responses to clicks as a function of sound intensity in the guinea pig.

Hearing Research 59(2):138–156. https://doi.org/10.1016/0378-5955(92)90111-Y

[29] Bruce IC, Erfani Y, Zilany MS (2018) A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. Hearing Research 360:40–54. https://doi.org/10.1016/j.heares.2017.12.016

[30] Kiang NYS, Moxon EC, Kahn AR (1976) The Relationship of Gross Potential Record from the Cochlea to Single Unit Activity in the Auditory Nerve. In: Electrocochleography. University Park Press Baltimore, p 95–115

[31] Wang B (1979) The relation between the compound action potential and unit discharges of the auditory nerve. PhD thesis, MIT

[32] Prijs VF (1986) Single-unit response at the round window of the guinea pig. Hearing Research 21(2):127–133. https://doi.org/10.1016/0378-5955(86)90034-1

[33] Deloche F (2022) fdeloche/fmaskedCAP-model: v1.0. https://doi.org/10.5281/ZENODO.6403024, URL https://doi.org/10.5281/zenodo.6403023

[34] Paszke A, Gross S, Massa F, et al (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, et al (eds) Advances in Neural Information Processing Systems 32. Curran Associates, Inc., p 8024–8035

[35] Özdamar Ö, Dallos P (1978) Synchronous responses of the primary auditory fibers to the onset of tone burst and their relation to compound action potentials. Brain Research 155(1):169–175. https://doi.org/10.1016/0006-8993(78)90320-7

[36] Kiang NYS, Watanabe T, Thomas EC, et al (1965) Discharge patterns of single fibers in the cat's auditory nerve, vol 35. MIT press Cambridge, MA

[37] Ruggero MA (1992) Physiology and coding of sound in the auditory nerve. The mammalian auditory pathway: Neurophysiology pp 34–93

[38] McMahon CM, Patuzzi RB (2002) The origin of the 900 Hz spectral peak in spontaneous and sound-evoked round-window electrical activity. Hearing Research 173(1-2):134–152. https://doi.org/10.1016/S0378-5955(02)00281-2

[39] Chertoff ME, Earl BR, Diaz FJ, et al (2012) Analysis of the cochlear microphonic to a low-frequency tone embedded in filtered noise. The Journal of the Acoustical Society of America 132(5):3351–3362. https:

//doi.org/10.1121/1.4757746

[40] Parida S, Heinz MG (2022) Distorted Tonotopy Severely Degrades Neural Representations of Connected Speech in Noise following Acoustic Trauma. Journal of Neuroscience 42(8):1477–1490. https://doi.org/10.1523/JNEUROSCI.1268-21.2021

[41] Temchin AN, Rich NC, Ruggero MA (2008) Threshold tuning curves of chinchilla auditory nerve fibers. II. Dependence on spontaneous activity and relation to cochlear nonlinearity. Journal of Neurophysiology 100(5):2899–2906. https://doi.org/10.1152/JN.90639.2008

[42] Unoki M, Irino T, Glasberg B, et al (2006) Comparison of the roex and gammachirp filters as representations of the auditory filter. The Journal of the Acoustical Society of America 120(3):1474–1492. https://doi.org/10.1121/1.2228539

[43] Irino T, Patterson RD (2002) A time-domain, level-dependent auditory filter: The gammachirp. The Journal of the Acoustical Society of America 101(1):412–419. https://doi.org/10.1121/1.417975

[44] Lopez-Poveda EA, Meddis R (2001) A human nonlinear cochlear filterbank. The Journal of the Acoustical Society of America 110(6):3107–3118. https://doi.org/10.1121/1.1416197

[45] Simpson MJ, Jennings SG, Margolis RH (2020) Techniques for Obtaining High-quality Recordings in Electrocochleography. Frontiers in Systems Neuroscience 14 https://doi.org/10.3389/fnsys.2020.00018

[46] Goodman SS, Lichtenhan JT, Jennings SG (2023) Minimum Detectable Differences in Electrocochleography Measurements: Bayesian-Based Predictions. Journal of the Association for Research in Otolaryngology 24(2):217–237. https://doi.org/10.1007/s10162-023-00888-0

[47] Møller AR, Jannetta PJ (1981) Compound action potentials recorded intracranially from the auditory nerve in man. Experimental Neurology 74(3):862–874. https://doi.org/10.1016/0014-4886(81)90259-4,

[48] Møller AR, Jho HD (1989) Response from the exposed intracranial human auditory nerve to low-frequency tones: Basic characteristics. Hearing Research 38(1-2):163–175. https://doi.org/10.1016/0378-5955(89)90137-8,

[49] Huet A, Batrel C, Dubernard X, et al (2022) Peristimulus Time Responses Predict Adaptation and Spontaneous Firing of Auditory-Nerve Fibers: From Rodents Data to Humans. Journal of Neuroscience 42(11):2253–2267

[50] Alamri Y, Jennings SG (2023) Computational modeling of the human compound action potential. The Journal of the Acoustical Society of America 153(4):2376. https://doi.org/10.1121/10.0017863