A neural-network framework for modelling auditory sensory cells and synapses

Fotios Drakopoulos, Deepak Baby, Sarah Verhulst

Dept. of Information Technology, Ghent University, 9000 Ghent, Belgium

E-mail: * fotios.drakopoulos@ugent.be; deepakbabycet@gmail.com; s.verhulst@ugent.be

Abstract

In classical computational neuroscience, transfer functions are derived from neuronal recordings to derive analytical model descriptions of neuronal processes. This approach has resulted in a variety of Hodgkin-Huxley-type neuronal models, or multi-compartment synapse models, that accurately mimic neuronal recordings and have transformed the neuroscience field. However, these analytical models are typically slow to compute due to their inclusion of dynamic and nonlinear properties of the underlying biological system. This drawback limits the extent to which these models can be integrated within large-scale neuronal simulation frameworks and hinders an uptake by the neuro-engineering field which requires fast and efficient model descriptions. To bridge this translational gap, we present a hybrid, machine-learning and computational-neuroscience approach that transforms analytical sensory neuron and synapse models into artificial-neural-network (ANN) neuronal units with the same biophysical properties. Our ANN-model architecture comprises parallel and differentiable equations that can be used for backpropagation in neuro-engineering applications, and offers a simulation run-time improvement factor of 70 and 280 on CPU or GPU systems respectively. We focussed our development on auditory sensory neurons and synapses, and show that our ANN-model architecture generalizes well to a variety of existing analytical models of different complexity. The model training and development approach we present can easily be modified for other neuron and synapse types to accelerate the development of large-scale brain networks and to open up avenues for ANN-based treatments of the pathological system.

1 Introduction

Following the fundamental work of Hodgkin and Huxley in modelling action-potential generation and propagation [1], numerous specific neuronal models were developed that proved essential for shaping and driving modern-day neuroscience [2]. In classical computational neuroscience, transfer functions between stimulation and recorded neural activity are derived and approximated analytically. This approach resulted in a variety of stimulus-driven models of neuronal firing and was successful in describing the nonlinear and adaptation properties of sensory systems [3–6]. For example, the mechano-electrical transduction of cochlear inner-hair-cells (IHCs) was described using conductance models [7–10], and the IHC-synapse firing rate using multi-compartment diffusion models [11–13]. Such mechanistic models have substantially improved our understanding of how individual neurons function, but even the most basic models use coupled sets of ordinary differential equations (ODEs) in their descriptions. This computational complexity hinders their

further development to simulate more complex behaviour, limits their integration within large-scale neuronal simulation platforms, such as brain-machine interfaces [14, 15], and their uptake in neuro-engineering applications that require real-time, closed-loop neuron model units [16, 17].

To meet this demand, neuroscience recently embraced deep-learning [18]; a technique that quickly revolutionised our ability to construct large-scale neuronal networks and to quantify complex neuronal behaviour [19–28]. These machine-learning methods can yield efficient, end-to-end descriptions of neuronal transfer functions, population responses or neuro-imaging data without having to rely on detailed analytical descriptions of the individual neurons responsible for this behaviour. Deep Neural Networks (DNNs) learn to map input to output representations and are composed of multiple layers with simplified units that loosely mimic the integration and activation properties of real neurons [29]. Examples include DNN-based models that were succesfully trained to mimic the representational transformations of sensory input [30, 31], or DNNs that use neural activity to manipulate sensory stimuli [32, 33]. Even though deep-learning has become a powerful research tool to help interpret the ever-growing pool of neuroscience and neuro-imaging recordings [34, 35], these models have an important drawback when it comes to predicting responses to novel inputs. DNNs suffer from their data-driven nature that requires a vast amount of data to accurately describe an unknown system, and can essentially be only as good as the data that were used for training. Insufficient experimental data can easily lead to overfitted models that describe the biophysical systems poorly while following artifacts or noise present in the recordings [36]. The boundaries of experimental neuroscience and limited experiment duration hence pose a serious constraint on the ultimate success of ANN models in predicting responses of neuronal systems.

To overcome these difficulties and merge the advantages of analytical and ANN model descriptions, we propose a hybrid approach in which analytical neuronal models are used to generate a sufficiently large and diverse dataset to train DNN-based models of sensory-cells and synapses. A combination of traditional and machine-learning approaches was recently adopted to optimise analytical model descriptions [37], but our method moves in the opposite direction and takes advantage of deep-learning benefits to develop convolutional neural network (CNN) models from mechanistic descriptions of neurons and synapses. We show here that the resulting CNN models can accurately simulate outcomes of traditional Hodgkin-Huxley neuronal models and synaptic diffusion models, but in a differentiable and computationally-efficient manner. The CNN-based model architecture is compatible with GPU computing and facilitates the integration of our model-units within large-scale, closed-loop or spiking neuronal networks. The most promising design feature relates to the backpropagation property, a mathematically-complex trait to achieve for nonlinear, coupled ODEs of traditional neural models. We will illustrate here how normal and pathological CNN models can be used in backpropagation to modify the sensory stimuli such to yield an optimised (near-normal) response of the pathological system.

We develop and test our hybrid approach on sensory neurons and synapses within the auditory system. The cochlea, or inner-ear, encodes sound via the inner hair cells (IHCs). IHCs sense the vibration of the basilar membrane in response to sound using their stereocilia, and translate this movement into receptor potential changes. By virtue of Ca^{2+}-driven exocytosis, glutamate is released to drive the synaptic transmission between the IHC and the innervated auditory-nerve fiber (ANF) synapses and neurons [38]. Experimentally extracted IHC parameters from in-vitro, whole-cell patch clamp measurements of the cellular structures and channel properties [39, 40] have led to different model descriptions of the nonlinear and frequency-dependent IHC transduction [10, 41–43]. Parameters for analytical IHC-ANF synapse models have mainly been derived from single-unit AN recordings to basic auditory stimuli in cats and small rodents [44–50]. Progressive insight into

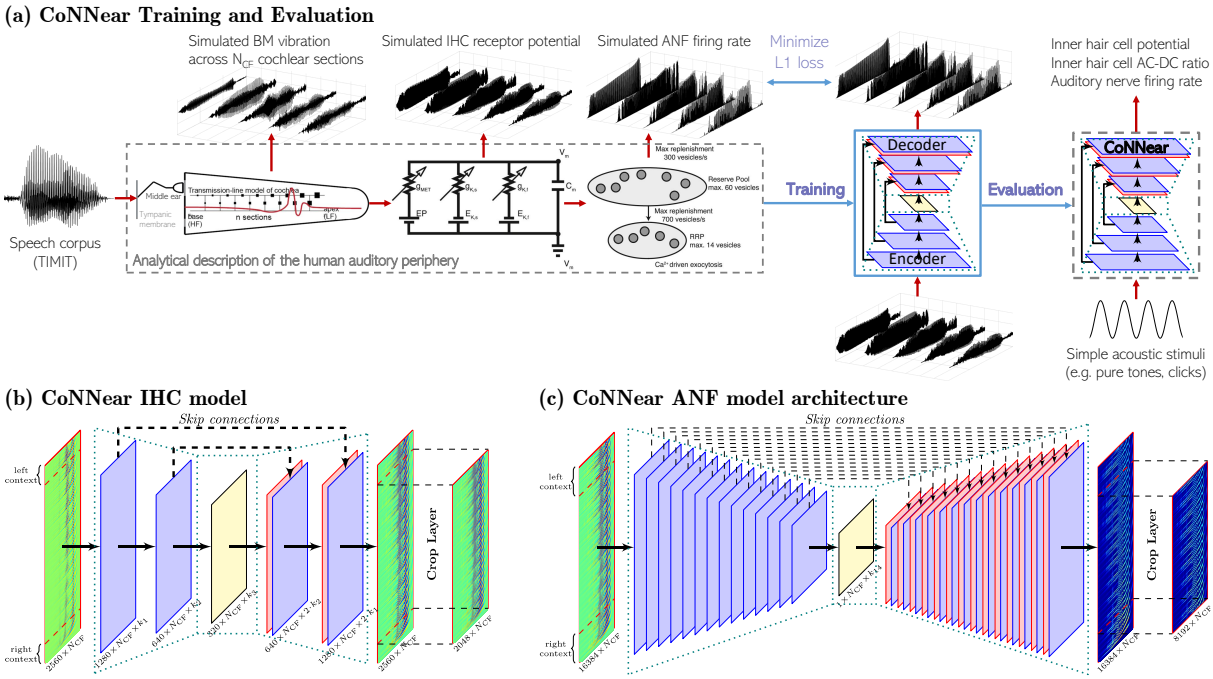

Fig 1. **(a)** Overview of the CoNNear model training and evaluation procedure. **(b)** Architecture of the CoNNear inner-hair-cell transduction model. **(c)** Generic architecture used for the CoNNear auditory-nerve-fiber synapse models.

the function of IHC-ANF synapses over the past decades has inspired numerous analytical model descriptions of the IHC, IHC-ANF synapse and ANF neuron complex [11–13, 51–61].

To generate sufficient training data for our CNN-based models of IHC-ANF processing, we adopted a state-of-the-art biophysical model of the human auditory periphery that simulates mechanical as well as neural processing of sound [60]. We describe here how the CNN model architecture and hyperparameters can be optimised for complex neuron or synapse models and we evaluate the quality of our CNN models on the basis of key IHC-ANF complex properties described in experimental studies, i.e., the IHC AC/DC ratio and excitation patterns, ANF firing rate, rate-level curves and modulation synchrony. The stimuli we adopted for the evaluations were not included in the CNN training datasets. We then determine the model run-time benefit over analytical models and investigate the extent to which our methodology is generalisable to different mechanistic descriptions of the IHC-ANF complex. Lastly, we provide two user cases: one in which IHC-ANF models are connected to a CNN-based cochlear mechanics model (CoNNear [62]) to capture the full transformation of acoustic stimuli into IHC receptor potentials and ANF firing rates along the cochlear tonotopy and hearing range, and a second one where we illustrate how backpropagation can be used to modify the CNN model input to restore a pathological output.

2 The CoNNear IHC and ANF models

Figure 1(a) depicts the adopted training and evaluation method to calibrate the parameters of each CoNNear module. Three modules that correspond to different stages of the analytical auditory periphery model described in Verhulst et al. [60] were considered: cochlear processing, IHC

transduction and ANF firing. The calibration of the cochlear mechanics module ($\text{CoNNear}_{\text{cochlea}}$) is described elsewhere [62, 63], here we focus on developing the sensory neuron models (i.e., $\text{CoNNear}_{\text{IHC}}$ and $\text{CoNNear}_{\text{ANF}}$). Fig. 1(a) illustrates the training procedure for the $\text{CoNNear}_{\text{ANf}_\text{L}}$ module that approximates the functioning of a low-spontaneous-rate ANF. Acoustic speech material is given as input to an analytical description of cochlear and IHC-ANF processing, after which simulated ANF firing rates are used as training material to determine the $\text{CoNNear}_{\text{ANf}_\text{L}}$ parameters. CoNNear modules were trained separately for each stage of the IHC-ANF complex, resulting in one model for IHC transduction and three models for different ANF types: a high- (H; 68.5 spikes/s), medium- (M; 10 spikes/s) and low- (L; 1 spike/s) spontaneous-rate (SR) ANF fiber. We chose a modular approach because this facilitates future simulations of the pathological system, where the IHC receptor potential can be impaired through presbycusis [64], or where selective damage to the ANF population can be introduced through cochlear synaptopathy [65].

Each module was modelled using a convolutional encoder-decoder architecture, consisting of a distinct number of CNN layers, as shown in Figs. 1(b),(c). Within these architectures, each CNN layer is comprised of a set of filterbanks followed by a nonlinear operation [18], except for the last layer where the nonlinear operation was omitted. The parameters of the nonlinear operations were shared across the frequency and time dimensions (first two dimensions) of the model, i.e., weights were applied only on the filter dimension (third dimension). The encoder CNN layers use strided convolutions, i.e., the filters were shifted by a time-step of two such to half the temporal dimension after every CNN layer. Thus, after N encoder CNN layers, the input signal was encoded into a representation of size $L/2^N \times k_N$, where k_N equals the number of filters in the N^{th} CNN layer. The decoder uses N deconvolution, or transposed-convolutional, layers, to double the temporal dimension after every layer to re-obtain the original temporal dimension of the input (L). Skip connections were used to bypass temporal audio information from the encoder to the decoder layers to preserve the stimulus phase information across the architecture. Skip connections have earlier been adopted for speech enhancement applications to avoid the loss of temporal information through the encoder compression [66–69], and could benefit the model-training to best simulate the nonlinear and the level-dependent properties of auditory processing by providing interconnections between several CNN layers [62, 70]. Lastly, we provided context information by making a number of previous and following input samples also available to the CoNNear modules when simulating an input of length L. Because CNN models treat each input independently, providing context is essential to avoid discontinuities at the simulation boundaries and take into account the neural adaptation processes [62]. A final cropping layer was added to remove the context after the last CNN decoder layer. Even though we trained each module using a fixed L, CoNNear models can process input of any length L and N_{CF} once they are trained due to their convolutional architecture.

To provide realistic input to the IHC-ANF models for training, acoustic speech waveforms were input to the cochlear model after which simulated cochlear basilar-membrane (BM) outputs were used to train and evaluate the IHC-ANF models. To this end, the IHC transduction model was trained using $N_{\text{CF}} = 201$ cochlear filter outputs that span the human hearing range (0.1-12kHz) and that were spaced according to the Greenwood place-frequency description of the human cochlea [71]. Similarly, simulated IHC receptor potentials of the analytical model cochlear regions ($N_{\text{CF}} = 201$) were used as training material for the different ANF models. The CoNNear model parameters of each module were optimised to minimise the mean absolute error (L1-loss) between the predicted CoNNear outputs and the reference analytical model outputs. It should be noted that even though we trained the models on the basis of 201 inputs, the optimal weights for a single CF-independent IHC or ANF model were determined during the training phase. Thus,

these model units can afterwards be connected to each N_{CF} input to simulate CF-dependent IHC or ANF processing of the entire cochlea.

To evaluate the CoNNear IHC-ANF models, it is important to characterise their properties to acoustic stimuli that were not seen during training. Training was performed using a single speech corpus [72], but IHC and ANF processing have very distinct adaptation, and frequency- and level-dependent properties to basic auditory stimuli such as tones, clicks or noise. To test how well the CoNNear modules generalise to unseen stimuli and to other analytical IHC-ANF model descriptions, we evaluated their performance on a set of classical experimental neuroscience recordings of IHC transduction and ANF firing. The six considered evaluation metrics together form a thorough evaluation of the CoNNear IHC-ANF complex, and outcomes of these simulations were used to optimise the final model architecture and its hyperparameters. Additional details on the model architecture, training procedure and IHC-ANF evaluation metrics are given in Methods.

In the following sections, we first describe how we optimised the architectures of each CoNNear model. We evaluate how well the trained CoNNear models capture signature experimental IHC and ANF processing properties using stimuli that were not part of the model training. Afterwards, we quantify the run-time benefit of our CoNNear models over the analytical descriptions and show how the modules can be connected to simulate processing of the entire cochlear periphery. To demonstrate the versatility of our method, we describe the extent to which our methodology can be applied to different mechanistic descriptions of the IHC-ANF complex. And lastly, we illustrate how the differential properties of our CoNNear models can be used within a closed-loop backpropagation network to restore the function of a pathological system.

3 Determining the CoNNear hyperparameters

Table 1 shows the final layouts of all the CoNNear modules we obtained after taking into account: (i) the L1-loss on the training speech material (i.e., the absolute difference between simulated CNN and analytical responses), (ii) the desired auditory processing characteristics, and (iii) the computational load. A principled fine-tuning approach was followed for each CoNNear module architecture on which we elaborate in the following sections.

3.1 CoNNear IHC model

Fixed parameters: Prior knowledge of fine-tuning a neural-network-based model of human cochlear processing [62] helped us to make initial assumptions about the needed architecture to accurately capture the computations performed by the analytical IHC model [60]. The shorter IHC adaptation time constants [73] enabled us to use fewer convolution layers and shorter filter lengths than were used in the CoNNear cochlea model. Thus, we opted for an architecture with 6 convolution layers and a filter length of 16. In each layer, 128 convolution filters were used and a stride of 2 was applied for dimensionality reduction. Each layer was followed by a hyperbolic-tangent (tanh) nonlinearity. The input length was set to $L_c = 2048 + 2 \cdot 256 = 2560$ samples (102.8 ms). Figure 2 shows that the trained architecture (b) generally followed the pure-tone excitation patterns of the reference model (a), but showed a rather noisy response across CFs, especially for the higher stimulation levels. Although we tried architectures with different layer numbers or filter durations, these models did not show significant improvements over Fig. 2(b) and usually resulted in noisier responses, or responses with a smaller dynamic range.

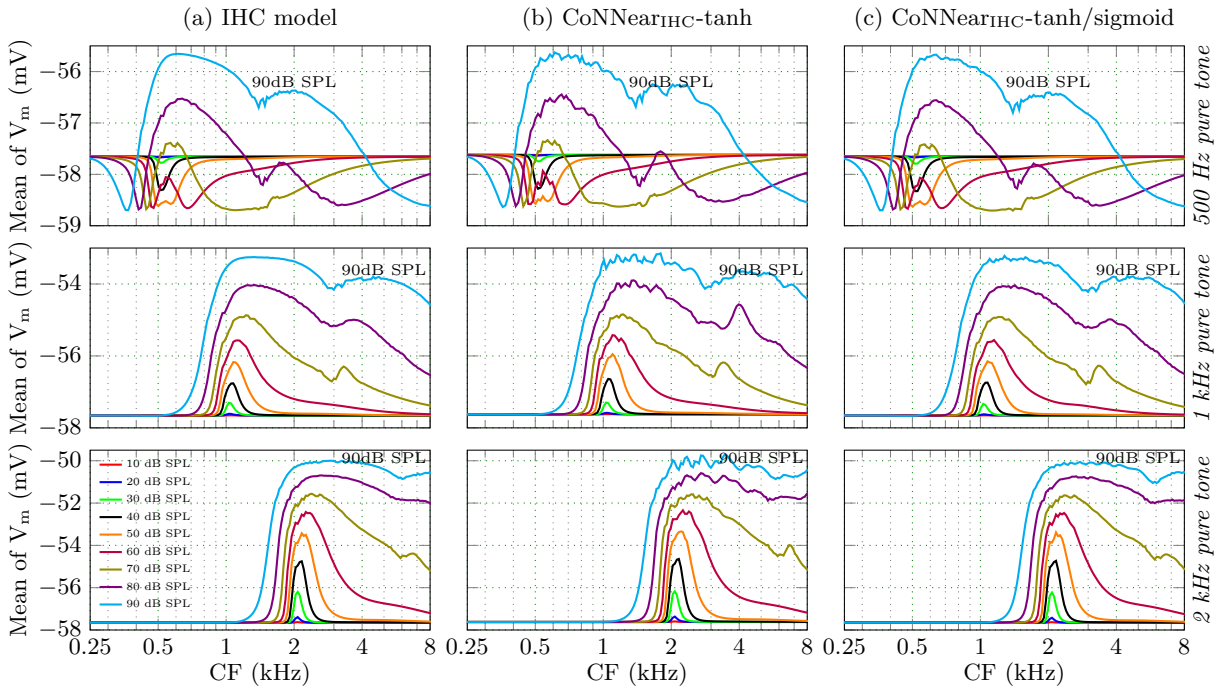

Fig 2. Comparing IHC excitation patterns. Simulated average IHC receptor potentials across CF for tone stimuli, presented at levels between 10 and 90 dB SPL. From top to bottom, the stimulus tone frequencies were 500Hz, 1 kHz and 2 kHz, respectively.

Hyperparameters: The shape of the activation function, or nonlinearity, is crucial to enable CoNNear to learn the level-dependent cochlear compressive growth properties and negative signal deflections present in BM and IHC processing. A *tanh* nonlinearity was initially preferred for each CNN layer, since it shows a compressive characteristic similar to the outer-hair-cell (OHC) input/output function [74] and crosses the x-axis. To optimise the rather noisy response of the trained IHC model (Fig. 2(b)) different nonlinear activation functions were compared for the encoder and decoder layers. Because the IHC receptor potential is expressed as a (negative) voltage difference, we opted for a *sigmoid* nonlinear function in the decoding layers to better capture the reference model outputs, while ensuring that the compressive nature present in the *tanh* could be preserved. Figure 2(c) shows that using a *sigmoid* activation function instead of a *tanh* for the decoder layers outperforms the *tanh* architecture (b) and better predicts the excitation patterns of the reference model (a).

Figure 3 furthermore depicts how different combinations of activation functions affected the simulated AC/DC ratios of the IHC responses across CF, and the half-wave rectified IHC receptor potential as a function of stimulus level. The logarithmic decrease of the AC/DC ratio and the linear-like growth of the IHC potential were predicted similarly using both architectures, but the *tanh* architecture overestimated the responses for high frequencies and levels. Overall, a much smoother response was achieved when using a *sigmoid* activation function in the decoder layers, motivating our final choice for the CoNNear IHC architecture (Table 1).

3.2 CoNNear ANF models

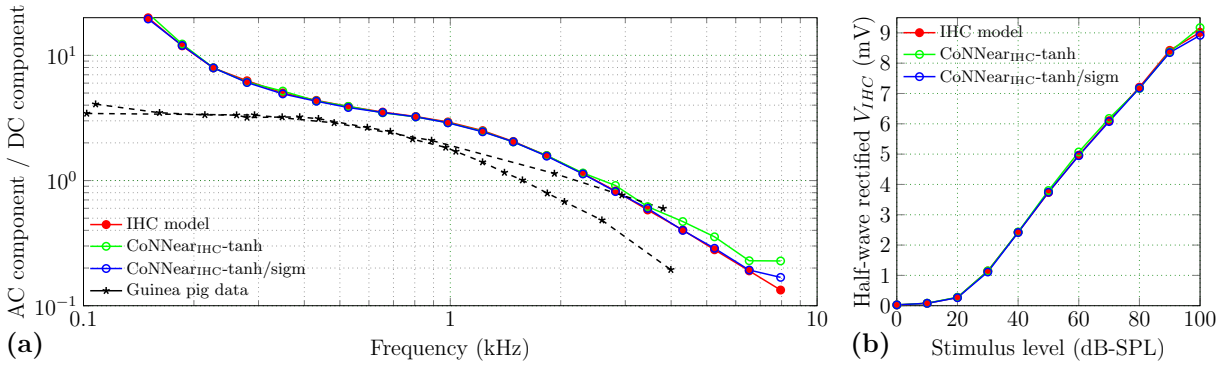

Fig 3. Comparing IHC transduction aspects. (a) Ratio of the AC and DC components of the IHC responses across CF for 80 dB SPL pure-tone bursts compared against guinea pig data [73]. **(b)** Root-mean-square of the half-wave rectified IHC receptor potential V_{IHC} in response to a 4-kHz pure-tone plotted as function of sound level.

Fixed parameters: Our modular approach enabled the use of the preceding stages to optimise our ANF model parameters. To determine the architecture, we first took into account the slower adaptation time constants (i.e., the much longer exponential decay) of the analytical ANF model description compared to those observed in the cochlea and IHC [12]. The choice of the window size L will thus be important to realistically capture the steady-state ANF response to sustained stimulation, e.g., for acoustic pure tones. Figure 4(a) visualises the exponential decay of simulated ANF firing rates, in response to a 1-kHz pure-tone presented at 70 dB SPL. At the time corresponding to a window size of 2048 samples (∼100 ms), the firing rates of the three ANFs have not significantly decayed to their steady state and hence we chose to use a longer window duration of L of 8192 samples (∼400 ms) for our ANF models. At 400 ms, the firing rates of the HSR, MSR and LSR fibers have respectively reached 99.5%, 95% and 93.4% of their final (1-sec) firing rate (Fig. 4(a)).

Another important aspect relates to capturing the experimentally [49] as well as computationally [60] observed slow recovery of ANF onset-peak responses after prior stimulation. Since CNN models treat each input independently, the duration of the context window is crucial to sufficiently present prior stimulation to the CoNNear ANF models. Figure 4(b) shows the exponential recovery of the onset-peak, for simulated responses of the three ANF types, as a function of the inter-stimulus interval between a pair of pure-tones. Two 2-kHz pure-tones were generated according to experimental procedures [49], i.e., 100 ms pure-tones presented at 40 dB above the firing threshold of each ANF (60, 65 and 75 dB for the HSR, MSR and LSR fibers respectively) with an inter-stimulus interval from 0.1 to 1.9 secs. Since the 1.9-sec interval corresponds to 38,000 samples (f_s = 20 kHz), a compromise was made to select a final context window that was short enough to limit the needed computational resources, but that was still able to capture the recovery and adaptation properties of the ANF models faithfully. We chose 7936 samples for the context window (∼400 ms) which resulted in a total input size of $L_c = 7936 + 8192 + 256 = 16384$ samples. For a 400-ms inter-stimulus interval, the onset-peak of the HSR, MSR and LSR fibers has recovered to the 92.4%, 94.2% and 95.8% of the onset-peak of the 1.9-sec interval tone respectively (Fig. 4(b)).

We further illustrate the effect of adding a context window in Fig. 5, by simulating responses of two trained CoNNear ANf$_{\text{L}}$ models to a 8192-sample-long 70-dB-SPL speech segment. Considering

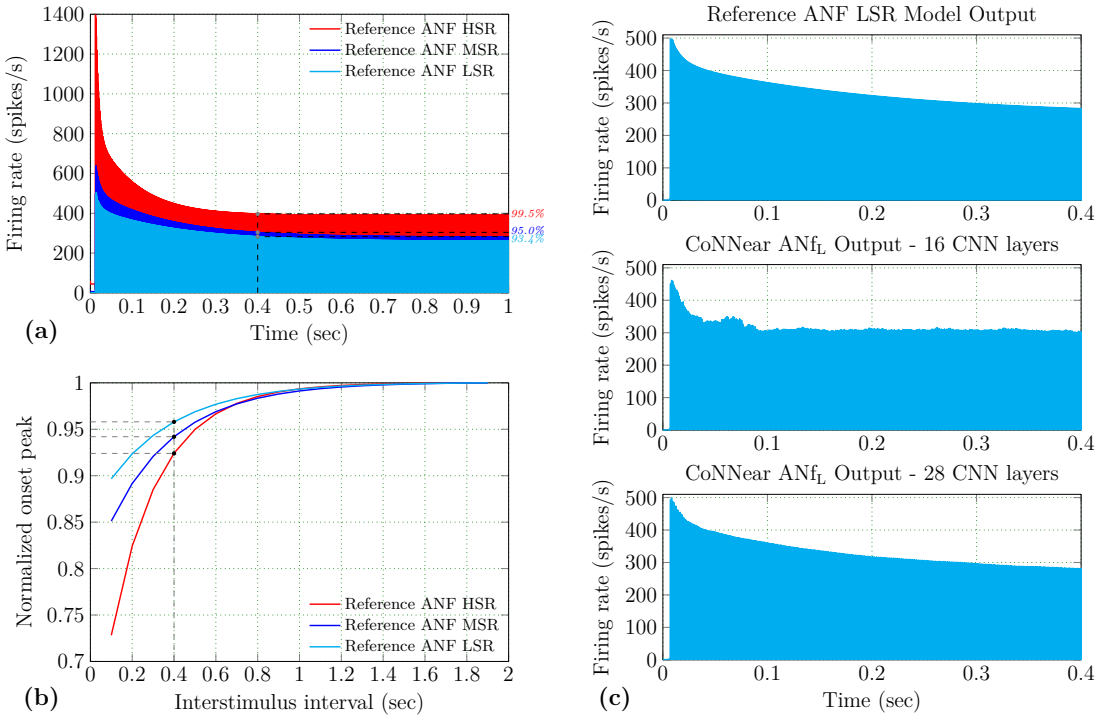

Fig 4. Parameter selection for the ANF models. (a) The firing rate of the three ANF models is shown over time, as a response to a 1-sec long 1 kHz pure-tone presented at 70 dB SPL. The percentages on the right side correspond to the percent of the steady state value that the firing rate has reached at 0.4 secs for each fiber. **(b)** The normalised amplitude of the onset peak is shown for a pair of 2 kHz pure-tones with interstimulus intervals from 0.1 to 1.9 seconds. Each time, the maximum value of the response to the second tone is reported, normalised by the maximum value of the response to the second tone with the longest interstimulus interval (1.9 sec). **(c)** From top to bottom, the simulated ANF LSR firing rate is shown for the reference ANF model, a trained model with 8 encoder layers and a trained model with 14 encoder layers, in response to a 70 dB pure-tone at 70 dB SPL.

an architecture with a short context window (c), the simulated response was unable to reach the onset amplitude of the reference LSR fiber model (b) observed for the high CFs at approximately 100 ms (grey dashed box). At the same time, the response for the short context architecture decayed to a more saturated output after the onset peak, compared to the reference model. In contrast, when adopting an architecture with a longer context window (d), the CoNNear ANf$_L$ model better captured the onset peak observed after the long inter-stimulus interval while showing an unsaturated fiber response which was similar to the reference model (b).

Lastly, the much slower adaptation properties of the ANF responses and the chosen input size of $L_c = 16384$ samples led us to realise that a larger number of CNN layers might be required to model the ANF stage, compared to the IHC transduction stage. A much deeper architecture might be necessary to simultaneously capture the characteristic ANF onset-peak and subsequent exponentially-decaying adaptation properties of ANF firing rates to step-like stimuli, and this is demonstrated in Fig. 4(c). A trained architecture consisting of 16 layers failed to capture the adaptation properties of the LSR ANF, while the use of 28 layers successfully approximated the

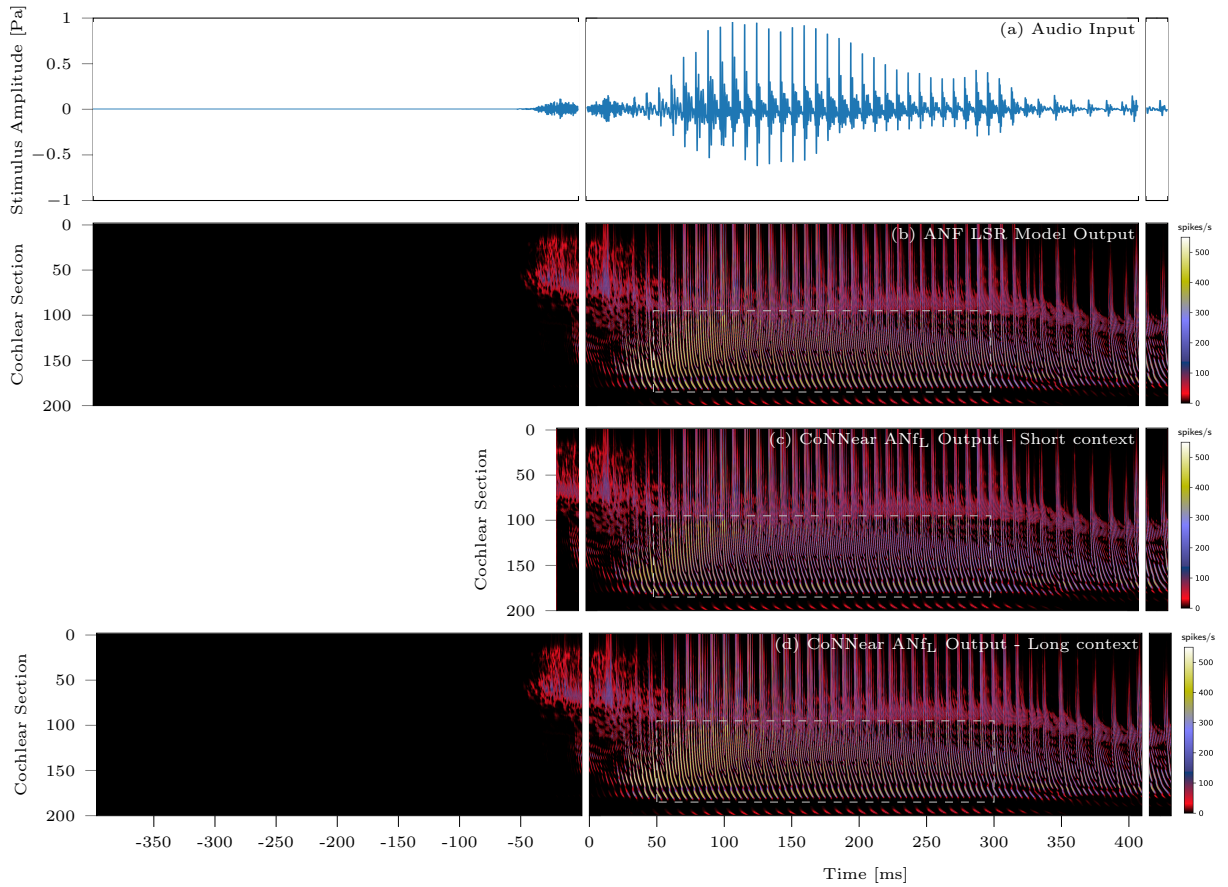

Fig 5. Simulated ANF firing rates for a 8192-sample-long speech stimulus. The stimulus waveform is depicted in panel (a) and panels (b)-(d) depict the output firing rate (in spikes/s) of the reference ANF LSR model (b) and two CoNNear ANF LSR architectures, with a context of 256 samples (c) and 7936 samples (d) included on the left side of the input respectively. The audio stimulus was presented to the reference cochlear and IHC model and the simulated IHC receptor potential output was used to stimulate the three ANF models. The N_{CF}=201 considered output channels are labeled per channel number: channel 0 corresponds to a CF of 100 Hz and channel 200 to a CF of 12 kHz.

reference firing rates. By introducing an architecture which encodes the input to a very condensed representation, preferably of a size of 1, we can ensure that the long-term correlations existent in the input can be captured faithfully by the convolutional filters. To this end, we opted for an architecture of 28 total layers and a condensed representation of size $1 \times N_{CF}$ (Fig. 1(c)) when using a stride of 2 and an input size of $L_c = 8192 + 7936 + 256 = 16384$ samples (819.2 ms). Since we adopted a deep architecture for the ANF models, the use of a long filter length was superfluous. Hence, we decreased it from 16 to 8 when compared to the IHC model. The lower number of coupled ODEs in the analytical ANF model description led us to further decrease the filter size used in every layer from 128 to 64 filters per layer.

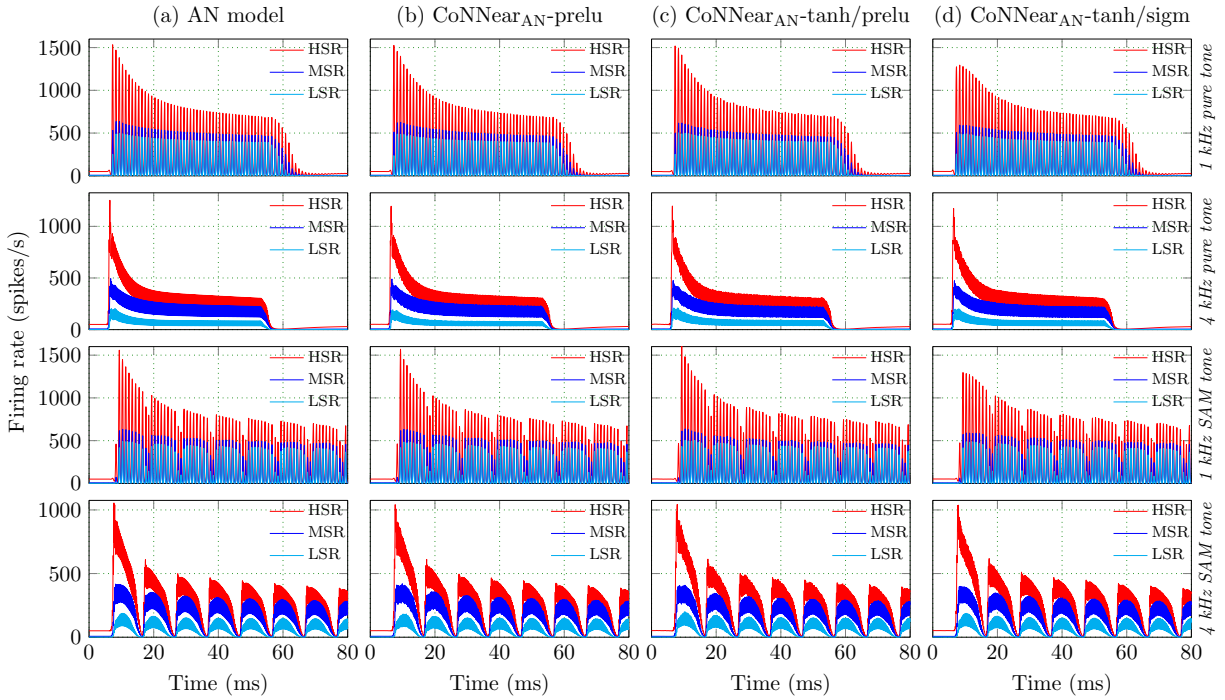

Fig 6. Comparing firing rates among the different ANF models. Simulated ANF firing rate across time for tone stimuli presented at 70 dB SPL. The blue, red and cyan graphs correspond to the responses of the HSR, MSR and LSR fiber models respectively. From top to bottom, the stimuli were 1 kHz, 4 kHz pure tones and 1kHz, 4 kHz amplitude-modulated tones.

Hyperparameters: The compressive properties of BM and IHC processing are not retained in ANF processing, so a linear activation function (a Parametric ReLU; *PReLU*) was initially used for each CNN layer. Figure 6 shows the responses of the three trained CoNNear ANF models (b) for different tonal stimuli in comparison to the reference ANF model (a). The firing rates of the three ANF models, CoNNear$_{\mathrm{ANf_H}}$, CoNNear$_{\mathrm{ANf_M}}$ and CoNNear$_{\mathrm{ANf_L}}$, are visualised in red, blue and cyan respectively.

The good match between analytical and CoNNear predictions in Fig. 6 was extended to ANF rate-level growth as well (Fig. 7), and together these simulations show that the chosen architecture and *PReLU* non-linearity were suitable to model characteristic ANF properties of the three ANF types. Compared to the reference firing rates, the architectures in panel (b) introduced noise, which might be eliminated by using a more compressive activation function (*tanh*) between the encoder layers. The *tanh* function was able to transform the negative potential of the IHC stage to the positive firing response of the ANFs (Fig. 6(c)), and yielded similar firing rates for all ANF models. However, for the CoNNear$_{\mathrm{ANf_H}}$ and CoNNear$_{\mathrm{ANf_M}}$ architectures, the *tanh* non-linearity introduced an undesirable compressive behaviour at higher stimulus levels, as depicted in Fig. 7(a). This was not the case for CoNNear$_{\mathrm{ANf_L}}$, and hence we also tested using a *sigmoid* nonlinearity in the decoder layers. This combination of non-linearities (d) compressed the responses of the CoNNear$_{\mathrm{ANf_H}}$ and CoNNear$_{\mathrm{ANf_M}}$ models even more, and negatively affected the onset responses. However, this combination (d) was found to best approximate the *LSR* ANFs firing rates. Table 1 summarizes the final parameters we chose for each of CoNNear ANF architecture based on simulations in Figs. 6 and 7.

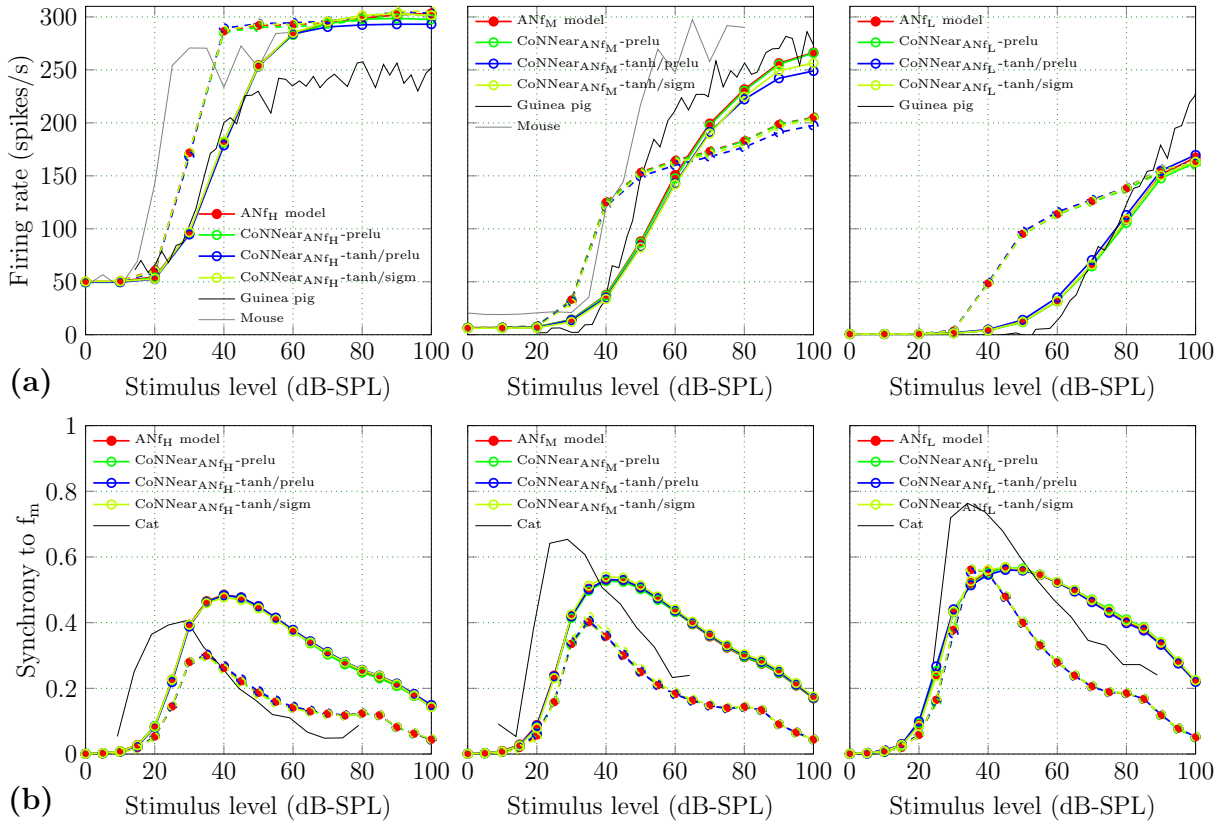

Fig 7. Level-dependent properties of the different ANF models. (a) From left to right, ANF rate-level curves were simulated for the HSR, MSR and LSR ANF models respectively, at CFs of 1 (dashed colored) and 4 kHz (solid colored). The reference data stemmed from guinea pig (fibers with SRs of 65 spikes/s, 10 spikes/s and 0 spikes/sec at a CF of ∼1.5 kHz; Fig. 1 in [47]) and mouse recordings (CF of 18.8 kHz for SR of 47.6 spikes/s and CF of 23.7 kHz for SR of 0.1 spikes/s; Fig. 6 in [75]). **(b)** From left to right, ANF synchrony-level functions were calculated for the HSR, MSR and LSR ANF models. For each ANF model, 1 kHz and 4 kHz pure tone carriers were modulated by an $f_m = 100$ Hz pure tone and presented at CFs of 1 (dashed colored) and 4 kHz (solid colored). For each CF, vector strength to the f_m is reported against the stimulus intensity for the three fiber types. The reference data came from cat ANF recordings (fibers of 8.1 kHz CF and 2.6 spikes/s, 1.14 kHz CF and 6.3 spikes/s, and 2.83 kHz and 73 spikes/s, respectively; Figs. 5 and 8 in [50]).

4 Evaluating simulated and recorded IHC-ANF properties

The excitation patterns of the final CoNNear IHC model (Fig. 2(c)) are generally consistent with the reference IHC model (a). The IHC AC/DC components (Fig. 3(a)) followed the simulated and measured curves well, and showed a slight overestimation for the lower frequency responses. The simulated half-wave-rectified IHC receptor potential (Fig. 3(a)) corroborated the in-vivo guinea pig IHC measurements [76], by showing an unsaturated, almost linear, growth of the half-wave rectified IHC receptor potential (in dB) for stimulation levels up to 90 dB.

The properties of single-unit ANF responses were accurately captured by the CoNNear

Table 1. Final parameter selection of the CoNNear architectures. The input length of each model is $L_c = L_l + L + L_r$ and the output length (after cropping) is L samples. The specified lengths L were used during training, but each architecture can process inputs of variable lengths L after training.

Parameters	L	L_l	L_r	Total Layers	Filters /Layer	Filter length	Encoder activation	Decoder activation
$\text{CoNNear}_{\text{IHC}}$	2048	256	256	6	128	16	tanh	sigmoid
$\text{CoNNear}_{\text{ANf}_\text{H}}$	8192	7936	256	28	64	8	PReLU	PReLU
$\text{CoNNear}_{\text{ANf}_\text{M}}$	8192	7936	256	28	64	8	PReLU	PReLU
$\text{CoNNear}_{\text{ANf}_\text{L}}$	8192	7936	256	28	64	8	tanh	sigmoid

architectures, as visualised in Figs. 6 and 7. For each ANF, the final architectures (Table 1) followed the reference model firing rates across time (Fig. 6). As expected, phase-locking to the stimulus fine-structure was present for the 1-kHz ANF response and absent for the 4-kHz ANF. Phase-locking differences between the 1 and 4-kHz CF fibers were also evident from their responses to amplitude-modulated tones.

The level-dependent properties of different ANF types were also captured by our CoNNear architectures, as shown in Fig. 7. Compared to the reference data, the 4-kHz simulations captured the qualitative differences between LSR, MSR and HSR guinea pig ANF rates well. The mouse rate-level curves show somewhat steeper growth than our simulations, especially when comparing the lower SR fiber data with the simulated MSR fiber responses. Given that the cochlear mechanics are fundamentally different across species, it is expected that the responses are not overly generalisable across species. The shape of the simulated rate-level curves was different for the different CF fibers (1-kHz dotted lines compared to 4-kHz solid lines) despite the CF-independent parameters of the ANF model. This illustrates that differences in BM processing across CF, given as input to the IHC-ANF model, are still reflected in the shape of ANF rate-level curves. The smaller dynamic range of levels encoded by the BM for the 1-kHz than the 4-kHz CF (e.g., Fig. 2 in [60]) was also reflected, yielding ANF level-curve compression at lower stimulus levels for the 1-kHz CF.

Lastly, ANF synchrony-level curves were overall captured well by our final CoNNear ANF architectures, while showing no apparent differences between the different non-linearities (Fig. 7(b)). In qualitative agreement with the reference data, the maxima of the synchrony-level curves shifted towards higher levels as the fibers' threshold and rate-level slope increased. At the same time, enhanced synchrony for LSR over HSR fibers was observed for medium to high stimulus levels, with the most pronounced difference for the 1-kHz simulations (dotted lines). For MSR and LSR fibers, the CoNNear models were able to simulate modulation gain, i.e., vector strength > 0.5 [50].

5 CoNNear as a real-time model for audio applications

The CoNNear IHC-ANF computations can be sped up when run on an AI accelerator (GPU, VPU etc.). Table 2 summarises the computation time required to execute the final CoNNear architectures on a CPU and GPU, for 201-channel and single-channel inputs. A TIMIT speech utterance of 2.3 secs was used for this evaluation and served as input to the analytical model [60] to simulate the outputs of the cochlear and IHC stages. The cochlear BM outputs were then framed into windows of 2560 samples (102.4 ms) to evaluate the CoNNear IHC model and the

Table 2. Model processing time. Comparison of the average time required to calculate each stage of the reference and the CoNNear model on a CPU (Intel Xeon E5-2620 v4) and a GPU (Nvidia GTX 1080). For each of the separate stages, the reported time corresponds to the average time needed to process a fixed-size input of $N_{CF} = 201$ frequency channels (population response) and $N_{CF} = 1$ channel (single-unit response), corresponding to the output of the preceding stage of the analytical model to a speech stimulus. The same results are shown for the CoNNear IHC-ANF complex, after connecting all the individual modules. The last row shows the computation time needed to transform a speech window input to ANF firing rates, after connecting the CoNNear cochlea and IHC-ANF modules together.

Model	**Trainable parameters**	**Window (samples)**	**CPU (s)**		**GPU (ms)**	
			201-CF	**1-CF**	**201-CF**	**1-CF**
IHC model	N/A	2560	1.2707	0.6117	N/A	
$\text{CoNNear}_{\text{IHC}}$	1,317,505	2560	1.0262	0.0102	56.40	2.18
ANf_{H} model	N/A	16384	1.0553	0.7197	N/A	
$\text{CoNNear}_{\text{ANf}_{\text{H}}}$	1,250,177	16384	2.6792	0.0289	178.25	7.21
ANf_{M} model	N/A	16384	1.0508	0.7015	N/A	
$\text{CoNNear}_{\text{ANf}_{\text{M}}}$	1,250,177	16384	2.6820	0.0279	175.97	6.95
ANf_{L} model	N/A	16384	1.0590	0.7019	N/A	
$\text{CoNNear}_{\text{ANf}_{\text{L}}}$	1,248,449	16384	2.2074	0.0243	115.86	4.53
IHC-ANF model	N/A	16384	9.7798	4.6532	N/A	
$\text{CoNNear}_{\text{IHC-ANF}}$	5,066,308	16384	11.8147	0.0676	803.48	16.61
Cochlea-IHC-ANF model	N/A	16384	167.4808	N/A	N/A	
$\text{CoNNear}_{\text{cochlea-IHC-ANF}}$	16,756,292	16384	12.6016	N/A	805.83	N/A

IHC outputs into windows of 16384 samples (819.2 ms) to evaluate the CoNNear ANF models. The average computation time is shown for each separate module of the IHC-ANF complex and the respective window size, as well as for the merged IHC-ANF model ($\text{CoNNear}_{\text{IHC-ANF}}$) after connecting all the separate modules together (see Methods). Lastly, our previously developed CoNNear cochlear model [62] was connected with $\text{CoNNear}_{\text{IHC-ANF}}$ to directly transform the speech inputs to ANF firing rates.

We did not observe a processing time benefit when running the IHC-ANF stages with 201-channel inputs on a CPU: the CoNNear ANF models actually increased the computation time on a CPU when compared to the reference models. However, the execution of the 201-channel IHC-ANF models on the GPU reduced the computation time 12-fold, when compared to the reference model CPU calculations. At the same time, our modular design choice makes it possible to use $\text{CoNNear}_{\text{IHC-ANF}}$ modules for only a subset of CFs or for single-unit responses. A significant speed up was seen for the latter case, with an almost 70-times faster CPU computation than for the reference model and a 280-times speed up when executed on the GPU. ANF firing rates can thus be simulated in ∼800 ms on a CPU, and less than 20 ms on a GPU for a stimulus window of more than 800 ms.

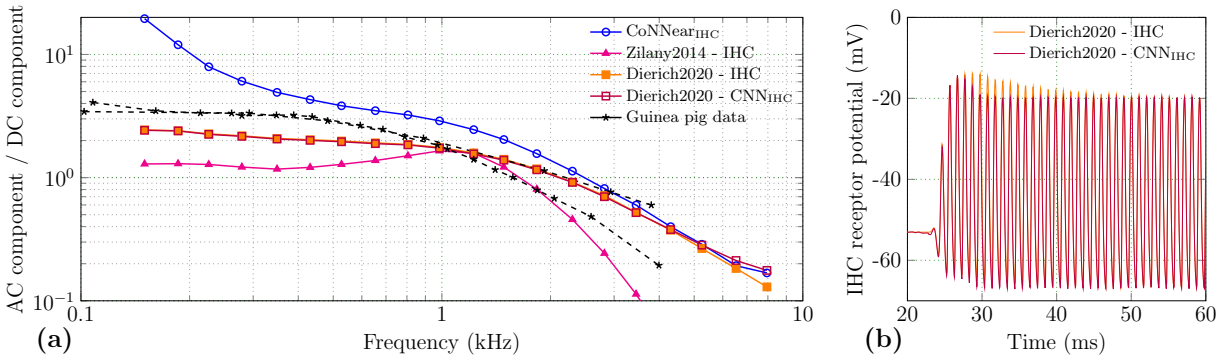

Fig 8. Comparison of different IHC analytical descriptions. (a) The ratio of the AC and DC components of the IHC responses is compared between different IHC analytical models across CF for 80-dB-SPL pure-tone bursts. **(b)** The IHC receptor potential output of the CNN approximation is compared against the baseline Dierich et al. IHC model [10], in response to a 1-kHz pure-tone of 70 dB SPL.

6 Generalisability of the framework to other analytical model descriptions

Here, we test whether our CNN-based IHC-ANF architectures can be used to approximate other analytical model descriptions of auditory neurons and synapses. This attests to the generalisability of our method for analytical model descriptions with varied levels of complexity. In Fig. 8, simulated AC/DC ratios are compared between responses of CoNNear$_{IHC}$ and two other state-of-the-art IHC analytical descriptions, the Zilany et al. [57] and Dierich et al. [10] models. The tonal stimuli described in Methods were used as inputs to the Zilany et al. cochlea-IHC model, while their extracted cochlear responses were used for the Dierich et al. IHC model. To demonstrate that the presented neural-network framework is generalisable to different Hodgkin-Huxley and diffusion-store model desciptions, we applied the IHC training approach (presented in Methods) to the Dierich et al. model. In the same fashion, the cochlear outputs we used for training CoNNear$_{IHC}$ were now used as inputs to the present model, and the same CNN architecture (Table 1) was trained using the new datasets. Figure 8(a) shows that the trained CNN model was able to accurately simulate the steady-state responses of this detailed IHC description, as reflected by the AC/DC ratio. However, a property that was not fully captured by our architecture was the adaptation of the responses after the stimulus onset, as shown in Figure 8(b). Due to the higher number of non-linearities comprised in this analytical model (i.e., 7 conductance branches in the Hodgkin-Huxley model), the CNN architecture might need to be adapted to include an additional layer or longer filter durations to yield more accurate simulations.

Figure 9 compares the ANF rate-level and synchrony-level curves between the responses of CoNNear$_{ANF}$ and two other state-of-the-art ANF descriptions, included in the Zilany et al. [57] and Bruce et al. [58] models. For both models, the auditory stimuli described in Methods were used as inputs to the respective cochlea-IHC and ANF descriptions and the results were computed from the post-stimulus time histogram (PSTH) responses using 100 stimulus repetitions. Once again, we applied the training approach of the CoNNear$_{ANF}$ architectures (see Methods) to approximate the HSR, MSR and LSR fiber models present in the Zilany et al. AN description. The same speech sentences were used as inputs to the Zilany et al. cochlea-IHC module to extract the IHC

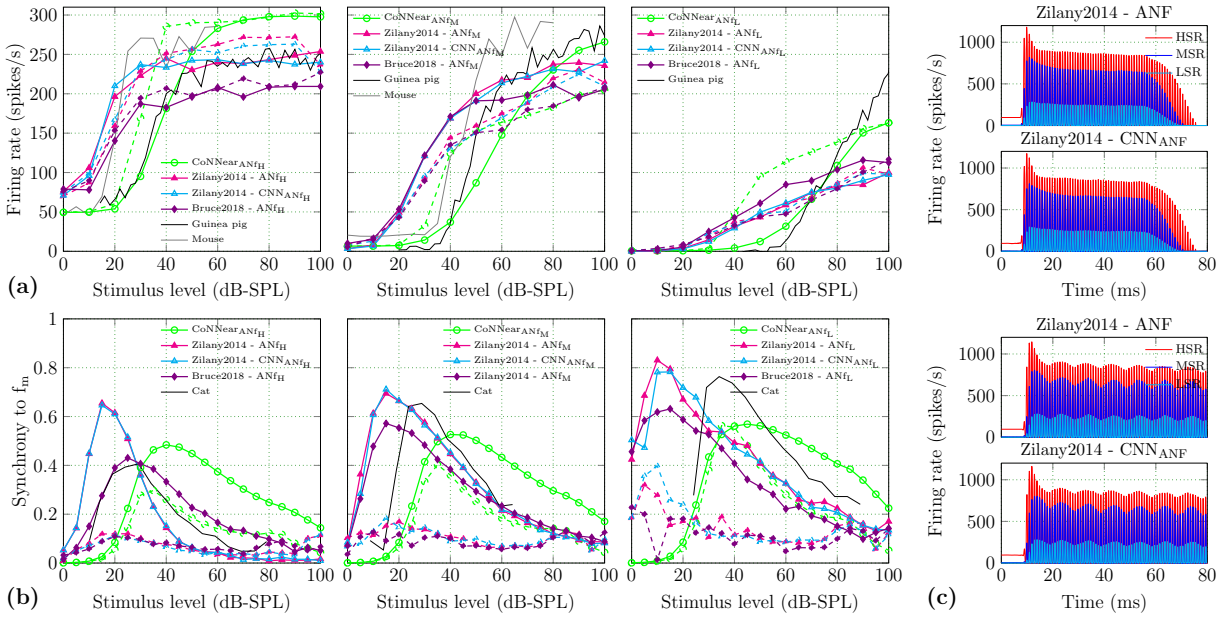

Fig 9. Comparison of different ANF analytical descriptions. (a) Rate-level curves for the HSR, MSR and LSR models of different ANF analytical descriptions, in response to tone-bursts at CFs of 1 (dashed colored) and 4 kHz (solid colored). **(b)** Synchrony-level functions for the HSR, MSR and LSR models of different ANF analytical descriptions, in response to modulated tones with carrier frequencies of 1 (dashed) and 4 kHz (solid) presented at CF. **(c)** For each fiber type, the ANF mean firing rate outputs of the CNN approximations are compared against the baseline Zilany et al. ANF model [57], in response to a 1-kHz tone-burst and a 1-kHz SAM tone of 70 dB SPL (f_m=100Hz).

potential responses, and the IHC outputs were subsequently given as inputs to the ANF module to extract the mean firing rates for each fiber type. The resulting datasets were used to train the CNN models, thus omitting the additive Gaussian noise and the subsequent spike generator due to the noisy and probabilistic character which is beyond the scope of our model architectures. However, after training the CNN models, the outputs can be fed to the spike generator present in the analytical ANF model to simulate PSTH responses.

The trained CNN models accurately approximated mean firing rates of the different ANF types, as shown in response to two different tonal stimuli (Fig. 9(c)). With the predicted outputs given as inputs to the spike generator model, the simulated PSTH responses were used to compute the ANF rate- and synchrony-level curves of the different types of ANFs (Fig. 9(a),(b)). The predicted curves show a similar trend to the Zilany et al. ANF model, however it is not possible to directly compare the resulting curves due to the inherent noise of the non-deterministic spike generator model in the reference analytical model.

7 CoNNear applications

Apart from the execution-time speed-up, an important benefit of CNN models over their respective analytical descriptions is given by their differentiable character. As a result, backpropagation algorithms can be computed from the outputs of these models to train new neural-networks. An

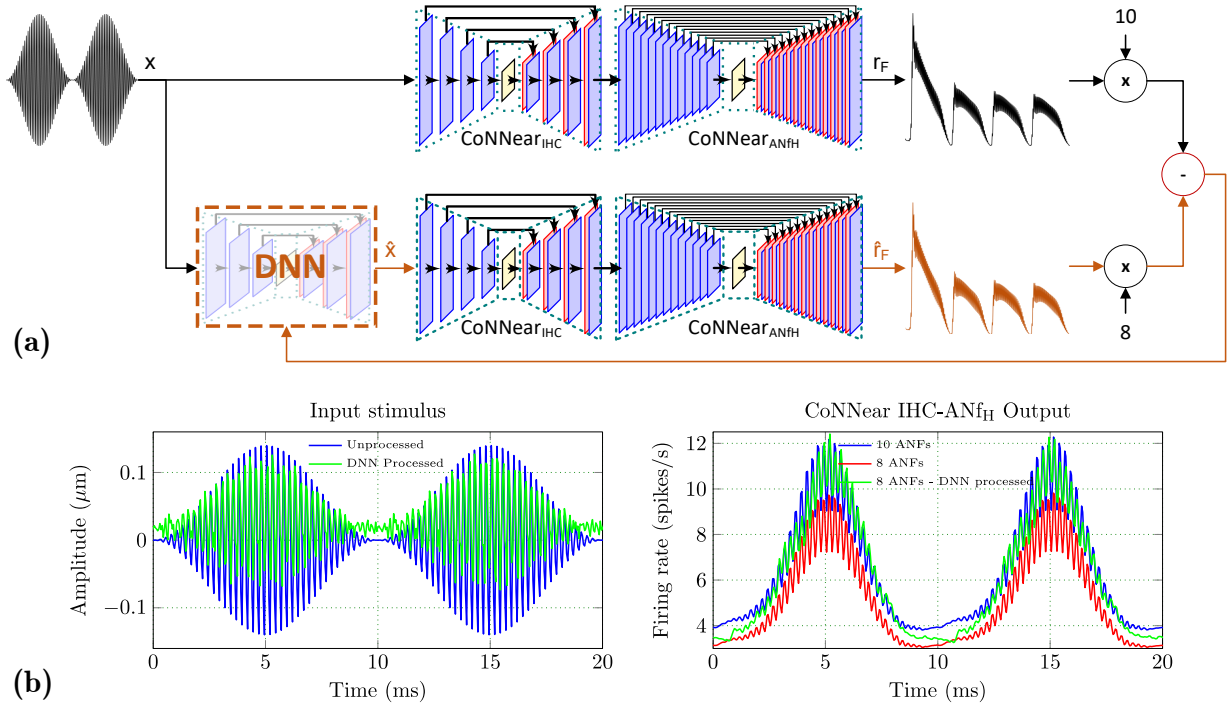

Fig 10. Training using CoNNear outputs. **(a)** The audio-signal processing DNN model is trained to minimise the difference of the outputs of the two CoNNear IHC-ANF models (orange pathway). **(b)** When processed by the trained DNN model, the input stimulus results to a firing rate output for the second model that closely matches the firing rate of the first model.

example user case is presented in Fig. 10(a), where a DNN model was trained to minimise the difference between the outputs of two IHC-ANF models: a normal and pathological model. Each model comprised the $\text{CoNNear}_{\text{IHC}}$ and $\text{CoNNear}_{\text{ANf}_\text{H}}$ modules, and the firing rates of each model were multiplied by a factor of 10 and 8 respectively, to simulate innervations of a normal-hearing human IHC at 4 kHz (Fig. 5 in [77]), and a pathological IHC that has a 20% fiber deafferentation due to cochlear synaptopathy [65]. The DNN model was trained based on the responses of these two CoNNear models to modify the stimulus such to restore the output of the pathological model back to the normal-hearing model output. Training was done using a small input dataset of 4 kHz tones with different levels and modulation depths, normalised to the amplitude ranges of IHC inputs, and the DNN model was trained to minimise the L1 loss between the time and frequency representations of the outputs. After training, the DNN model provides a processed input $\hat{x}$ to the 8-fiber model to generate an output $\hat{r}_\text{F}$ that matches the normal-hearing firing rate r_F as much as possible. The result for a modulated tone stimulus is shown in Fig. 10(b), for which the amplitude of the 8-fiber model response is restored to that of the normal-hearing IHC-ANF. This example demonstrates the backpropagation capabilites of our CNN models and their application range can be extended to more complex datasets such as a speech corpus, to derive suitable signal-processing strategies for speech processing restoration in hearing-impaired cochleae.

8 Discussion

Analytical descriptions of IHC-ANF processing have evolved over the years, with the IHC transduction shifting from simplified low-pass filter implementations [13, 51, 56, 57, 59, 78] to detailed models of basolateral outward K^+ currents [7–10]. State-of-the-art IHC-ANF models estimate the vibrations of the IHC stereocilia based on the mechanical drive to the IHC and often describe the ANF spikes or instantaneous firing rate resulting from the depletion and replenishment of different neurotransmitter stores [57, 58, 60, 79]. While such sensory models have progressed to accurately capture the nonlinear properties of human hearing, they typically comprise "hand-constructed" mechanistic descriptions that incorporate coupled sets of ODEs to describe small neuronal systems.

We presented a hybrid framework to develop a DNN-based model of IHC-ANF auditory processing, $\text{CoNNear}_{\text{IHC-ANF}}$. Different from pre-existing IHC-ANF models, the CoNNear architectures are based on DNNs that are differentiable and computationally efficient to accelerate and facilitate future studies of complex neuronal systems and behaviours. Our general framework for modelling sensory-cells and synapses consists of the following steps: (i) Derive an analytical description of the biophysical system using available neuroscience recordings. (ii) Use this analytical description to generate a training dataset that contains a broad and representative set of sensory stimuli. (iii) Define a suitable DNN-based architecture and optimise its hyperparameters on the basis of the training dataset and its performance on known physiological characteristics. (iv) Train the architecture to predict the behaviour of the biophysical system and evaluate using unseen data. Apart from requiring an analytical description that accurately describes the system, we showed that a careful design of the DNN architecture and a broad range of sensory input stimuli are essential to derive a maximally generalisable model.

The resulting IHC-ANF complex models were trained to apply the same operations to each frequency channel, such that they can be used for either single-unit or population response simulations across the cochlear partition. Simulating all $N_{CF} = 201$ frequency channels on the same CPU negatively impacted the required computation time compared to analytical descriptions, but nevertheless resulted in a biophysically-plausible, and rather versatile, model description. Single-channel $\text{CoNNear}_{\text{IHC-ANF}}$ CPU simulations did offer a 70-fold speed-up compared to their analytical counterparts, and can be executed with latencies below 20 ms when simulating ∼800 ms inputs on a GPU. This holds promise for the future uptake of our models within audio-processing pipelines that require real-time processing capabilities. At the same time, our models offer a differentiable solution that can directly be used in closed-loop systems for auditory feature enhancement or augmented hearing.

The trained DNN models can be further optimised using normal or pathological experimental data via transfer learning [63], or can be retrained on the basis of large neural datasets, when these become available. Approximately 3 and 8 days were needed to train each CoNNear ANF model and IHC model, respectively. To improve these training durations, a different scaling of the datasets, or batch normalisation between the convolutional layers, could prove beneficial [80]. Lastly, when considering their use for real-time applications, where ANF adaptation and recovery properties may be of lesser importance, it is possible to further reduce the context and window sizes of the ANF CoNNear models and bring execution times below 10 ms. However, this will always result in saturated, steady-state responses, rendering the models blind to long inter-stimulus intervals and unable to fully capture recovery properties, as visualised in Figs. 4 and 5. A different approach could be the use of recurrent layers (e.g., LSTM) within the CoNNear architectures to capture the dependency to prior stimulation without requiring long context windows.

We demonstrated that the proposed framework and architectures generalises well to unseen stimuli as well as to other auditory sensory cell and synapse model descriptions, and this provides a promising outlook. On the one hand, our method might be applicable to other neuronal systems that depend on nonlinear and/or coupled ODEs (e.g., see also their application to cochlear mechanics descriptions [62]). On the other hand, the CNN model architectures can easily be retrained when improved analytical model descriptions become available. When properly benchmarked, CNN-based neuronal models can provide new tools for neuroscientists to explain complex neuronal mechanisms such as heterogenous neural activity, circuit connectivity or optimisation [34, 35].

9 Conclusion

CoNNear presents a new method for projecting complex mathematical descriptions of neuron and synapse models, while providing a differentiable solution and accelerated run-time. Our proposed framework was applied to different auditory Hodgkin-Huxley neuron and synapse models, providing a baseline methodology for approximating nonlinear biophysical models of sensory systems. The presented $\text{CoNNear}_{\text{IHC-ANF}}$ model can simulate single-unit responses, speeding up the IHC-ANF processing, or population responses across a number of simulated tonotopic locations (default $N_{CF} = 201$) when connected to a cochlear model, preferably the $\text{CoNNear}_{\text{cochlea}}$ [62].

The developed CoNNear models are suitable for implementation in data processing devices such as a cochlear implant to provide biophysically-accurate stimuli to the auditory nerve. The ANF responses could also be used to drive neural-network back-ends that simulate brainstem processing or even the generation of auditory evoked potentials, such as the auditory brainstem response [59, 60] or the compound action potential [81]. All the developed CoNNear architectures can easily be integrated as part of brain networks, neurosimulators, or closed-loop systems for auditory enhancement or neuronal-network based treatments of the pathological system. Further neural network models can be developed on the basis of the present framework to compose large-scale neuronal networks and advance our understanding of the underlying mechanisms of such systems, making use of the transformative ability of backpropagating through these large-scale systems. We think that this type of neural networks can provide a breakthrough to delve deeper into unknown systems of higher processing levels, such as the brainstem, midbrain and cortical pathway of the human auditory processing.

Acknowledgments

This work was supported by the European Research Council (ERC) under the Horizon 2020 Research and Innovation Programme (grant agreement No 678120 RobSpear).

Competing interests

A patent application (PCTEP2020065893) was filed by UGent on the basis of the research presented in this manuscript. Inventors on the application are Sarah Verhulst, Deepak Baby, Fotios Drakopoulos and Arthur Van Den Broucke.

Data availability

The source code of the auditory periphery model used for training is available via 10.5281/zenodo.3717431 or github/HearingTechnology/Verhulstetal2018Model, the TIMIT speech corpus used for training can be found online [72]. All figures in this paper can be reproduced using the trained CoNNear models. A supplementary zip file is included which contains the evaluation procedure of the trained CoNNear models.

Code availability

The code for running and evaluating the trained CoNNear models, including instructions of how to execute the code, is included together with this manuscript and will be made available as a GitHub repository upon acceptance of this paper. A non-commercial, academic UGent license applies.

Author contributions

Fotios Drakopoulos: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing: Original Draft, Visualization; **Deepak Baby**: Conceptualization, Methodology; **Sarah Verhulst**: Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Writing: Review, Editing

Methods

The procedure of Fig. 1(a) was used to train the CoNNear IHC and ANF modules using simulated responses of an analytical Hodgkin-Huxley-type IHC model [43] and a three-store diffusion model of the ANF synapse [9] respectively. We adopted the implementations described in [60] and found on 10.5281/zenodo.3717431.

Figure 1(b) depicts the CoNNear IHC encoder-decoder architecture we used: an input of size $L_c \times N_{\mathrm{CF}}$ cochlear BM waveforms is processed by an *encoder* (comprised of three CNN layers) which encodes the input signal into a condensed representation, after which the *decoder* layers map this representation onto $L \times N_{\mathrm{CF}}$ IHC receptor potential outputs, for $N_{\mathrm{CF}} = 201$ cochlear locations corresponding to the filters' centre frequencies. Context is provided by making the previous $L_l = 256$ and following $L_r = 256$ input samples also available to an input of length $L = 2048$, yielding a total input size of $L_c = L_l + L + L_r = 2560$ samples.

The three CoNNear ANF models follow an encoder-decoder architecture as depicted in Fig. 1(c): an IHC receptor potential input of size $L_c \times N_{\mathrm{CF}}$ is first processed by an *encoder* (comprised of $N = 14$ CNN layers) which encodes the IHC input signal into a condensed representation of size $1 \times k_N$ using strided convolutions, after which the *decoder*, using the same number of layers, maps this representation onto $L \times N_{\mathrm{CF}}$ ANF firing outputs corresponding to $N_{\mathrm{CF}} = 201$ cochlear centre frequencies. Context is provided by making the previous $L_l = 7936$ and following $L_r = 256$ input samples also available to an input of length $L = 8192$, yielding a total input size of $L_c = L_l + L + L_r = 16384$ samples.

Training the CoNNear IHC-ANF complex

The IHC-ANF models were trained using reference analytical BM or IHC model simulations [60] to 2310 randomly selected recordings from the TIMIT speech corpus [72], which contains a large amount of phonetically balanced sentences with sufficient acoustic diversity. The 2310 TIMIT sentences were upsampled to 100 kHz to solve the analytical model accurately [82]. The root-mean-square (RMS) energy of half the sentences was adjusted to 70 dB and 130 dB sound pressure level (SPL), respectively. These levels were chosen to ensure that the stimuli contained sufficiently high instantaneous intensities, necessary for the CoNNear models to capture the characteristic input-output and saturation properties of individual IHC [47] and ANFs [76].

BM displacements, IHC potentials and ANF firing rates were simulated across 1000 cochlear sections with CFs between 25 Hz and 20 kHz [60]. The corresponding 1000 y_{BM}, V_m and $ANf_{h/m/l}$ output waveforms were downsampled to 20 kHz and only 201 uniformly distributed CFs between 112 Hz and 12 kHz were selected to train the CoNNear models. Above 12 kHz, human hearing sensitivity becomes very poor [83], motivating the chosen upper limit of considered CFs. The simulated data were then transformed into a one-dimensional dataset of $2310 \times 201 = 464310$ different training sequences. This dimension reduction was necessary because the IHC and ANF models are assumed to have CF-independent parameters, whereas the simulated BM displacements have different impulse responses for different CFs, due to the cochlear mechanics [84]. Hence, parameters for a single IHC or ANF model ($N_{CF} = 1$) were determined during training, based on simulated CF-specific BM inputs and corresponding IHC, or ANF outputs from the same CF.

For each of the resulting 464310 training pairs, the simulated BM and IHC outputs were sliced into windows of 2048 samples with 50% overlap and 256 context samples for the IHC model. In the case of the ANF models, silence was also added before and after each sentence with duration of 0.5 and 1 s, respectively, to ensure that our models can accurately capture the recovery and adaptation properties of ANF firing rates. The resulting simulated IHC and ANF outputs were sliced into windows of 8192 samples with 50% overlap, using 7936 context samples before and 256 samples after each window.

A scaling of 10^6 was applied to the simulated BM displacement outputs before they were given as inputs to the CoNNear IHC model, expressing them in [μm] rather than in [m]. Similarly, the simulated IHC potential outputs were multiplied by a factor of 10, expressed in [dV] instead of [V], and a scaling of 10^{-2} was applied to the simulated ANF outputs, expressing them in [x100 spikes/s]. These scalings were necessary to enforce training of CoNNear with sufficiently high digital numbers, while retaining as much as possible the datasets' statistical mean close to 0 and standard deviation close to 1 to accelerate training [80]. For visual comparison between the original and CoNNear outputs, the values of the CoNNear models were scaled back to their original units in all following figures and analyses.

CoNNear model parameters were optimised to minimise the mean absolute error (L1-loss) between the predicted CoNNear outputs and the reference analytical model outputs. A learning rate of 0.0001 was used with an Adam optimiser [85] and the entire framework was developed using the Keras machine learning library [86] with a Tensorflow [87] back-end.

After completing the training phase, the IHC and ANF models were extrapolated to compute the responses across all 201 channels corresponding to the $N_{CF} = 201$ tonotopic centre frequencies located along the BM. The trained architectures were adjusted to apply the same calculated weights (acquired during training) to each of the N_{CF} channels of the input, providing an output with the same size, as shown in Fig. 1(c). In the same way, the trained models can easily simulate single-CF IHC responses, or be used for different numbers of channels or frequencies than those

we used in the cochlear model.

Evaluating the CoNNear IHC-ANF complex

Three IHC and three ANF evaluation metrics were used to determine the final model architecture and its hyperparameters, and to ensure that the trained models accurately captured auditory properties, did not overfit to the training data and can be generalised to new inputs. Even though any speech fragment can be seen as a combination of basic stimuli such as impulses and tones of varying levels and frequencies, the acoustic stimuli used for the evaluation can be considered as unseen to the models, as they were not explicitly present in the training material. The evaluation stimuli were sampled at 20 kHz and had a total duration of 128 ms (2560 samples) and 819.2 ms (16384 samples) for the CoNNear IHC model and the CoNNear ANF models, respectively. The first 256 samples of the IHC stimuli and 7936 samples of the ANF stimuli consisted of silence, to account for the respective context of the models. Each time, the evaluation stimuli were passed through the preceding processing stages of the analytical model to provide the necessary input for each CoNNear model, i.e., through the cochlear model for evaluating the CoNNear IHC model and through the cochlear and IHC models for evaluating the CoNNear ANF models.

IHC excitation patterns

Inner-hair-cell excitation patterns can be constructed from the mean IHC receptor potential at each measured CF in response to tonal stimuli of different levels. Similar to cochlear excitation patterns, IHC patterns show a characteristic half-octave basal-ward shift of their maxima as stimulus level increases [88]. These excitation patterns also reflect the nonlinear compressive growth of BM responses with level observed when stimulating the cochlea with a pure-tone which has the same frequency as the CF of the measurement site in the cochlea [89].

We calculated excitation patterns for all 201 simulated IHC receptor potentials in response to pure tones of 0.5, 1 and 2 kHz frequencies and levels between 10 and 90 dB SPL using:

$$\text{tone}(t) = p_0 \cdot \sqrt{2} \cdot 10^{L/20} \cdot \sin(2\pi f_{\text{tone}} t), \tag{1}$$

where $p_0 = 2 \times 10^{-5}$Pa, L corresponds to the desired RMS level in dB SPL and f_{tone} to the stimulus frequencies. The pure-tones were multiplied with Hanning-shaped 5-ms ramps to ensure gradual onsets and offsets.

IHC transduction aspects

Palmer and colleagues recorded intracellular receptor potentials from guinea-pig IHCs in response to 80-dB-SPL tones [73], and reported the ratio between the AC and DC response components as a function of stimulus frequency. The AC/DC ratio shows a smooth logarithmic decrease over frequency and is used as a metric to characterise synchronisation in IHCs, with higher ratios indicating more efficient phase-locking. Our simulations were conducted for 80-ms, 80-dB-SPL tone bursts of different frequencies presented at the respective CFs, and were compared against experimental AC/DC ratios reported for two guinea-pig IHCs. We used a longer stimulus than adopted during the the experimental procedures (50 ms), to ensure that the AC component would reach a steady-state response after the stimulus onset. A 5-ms rise and fall ramp was used for the stimuli, and the AC and DC components of the responses were computed within windows of 50-70 ms after and 5-15 ms before the stimulus onset, respectively.

Capturing the dynamics of outward IHC K^+ currents has an important role in shaping ANF response properties of the whole IHC-ANF complex [9, 10]. This feature of mechanical-to-electrical transduction compresses IHC responses dynamically and thereby extends the range of v_{BM} amplitudes that can be encoded by the IHC, as postulated in experimental and theoretical studies [8,39]. As the v_{BM} responses only show compressive growth up to levels of 80 dB SPL [60,62], the simulated half-wave rectified IHC receptor potential is expected to grow roughly linearly with SPL (in dB) for stimulus levels up to 90 dB SPL, thus extending the compressive growth range by 10 dB. To simulate the IHC receptor potential, tonal stimuli with a frequency of 4 kHz and levels from 0 to 100 dB SPL were generated, using the same parameters as before (80-ms duration, 5-ms rise/fall ramp). The responses were half-wave rectified by subtracting their DC component, and the root-mean-square of the rectified responses was computed for each level.

ANF firing rates

We evaluate key properties of simulated ANF responses to amplitude-modulated and pure tone stimuli for which single-unit reference ANF recordings are available. We simulated the firing rate for low-, medium- and high- SR fibers to 1 and 4-kHz tone-bursts and amplitude-modulated tones, presented at 70 dB SPL and calculated at the respective CFs. Based on physiological studies that describe phase-locking properties of the ANF [50, 90], stronger phase-locking to the stimulus fine structure is expected for the 1-kHz fiber response than for the 4-kHz, where the response is expected to follow the stimulus envelope after its onset. Similar differences are expected for the amplitude-modulated tone responses as well.

The pure-tone stimuli were generated according to Eq. 1 and the amplitude-modulated tone stimuli using:

$$\text{SAM-tone}(t) = [1 + m \cdot \cos(2\pi f_{\mathrm{mod}} t + \pi)] \cdot \sin(2\pi f_{\mathrm{tone}} t), \tag{2}$$

where $m = 100\%$ is the modulation depth, $f_{mod} = 100$ Hz the modulation frequency, and f_{tone} the stimulus frequency. Amplitude-modulated tones were multiplied with a 7.8-ms rise/fall ramp to ensure a gradual onset and offset. The stimulus levels L were adjusted using the reference pressure of $p_0 = 2 \times 10^{-5}$ Pa, to adjust their root-mean-square energy to the desired level.

ANF level-dependent properties

Rate-level curves can be computed to evaluate ANF responses to stimulus level changes, in agreement with experimental procedures [47, 75]. Using Eq. 1, we generated pure-tone stimuli (50-ms duration, 2.5-ms rise/fall ramp) with levels between 0 and 100 dB and frequencies of approximately 1 and 4 kHz, based on the corresponding CFs of the ANF models (1007 and 3972.7 Hz). The rate levels were derived by computing the average response 10-40 ms after the stimulus onset (i.e., excluding the initial and final 10 ms, where some spike intervals may include spontaneous discharge [75]). Data from the experimental studies are plotted alongside our simulations and reflect a variety of experimental ANF rate-level curves from different species and CFs.

Synchrony-level functions were simulated for fully-modulated 400-ms long pure tones with a modulation frequency f_m of 100 Hz [50] and carrier frequencies of 1007 and 3972.7 kHz (henceforth referred to as 1 and 4 kHz), generated using Eq. 2. Synchrony to the stimulus envelope was quantified using vector strength [91] and was calculated by extracting the magnitude of the f_m component from the Fourier spectrum of the fibers' firing rate. The f_m magnitude was normalised to the DC component (0 Hz) of the Fourier spectrum, corresponding to the average firing rate of

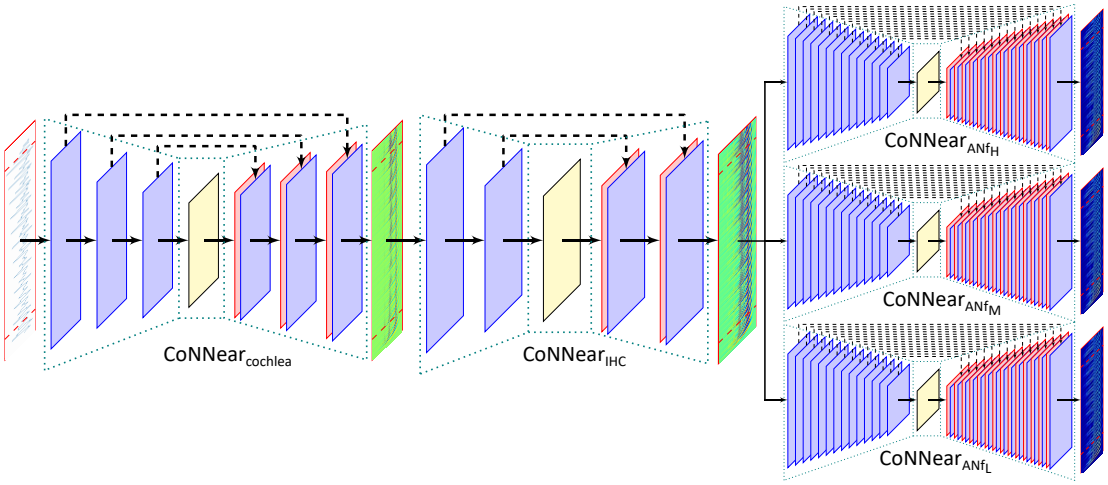

Fig 11. CoNNear model of the auditory periphery. Acoustic stimuli can be transformed to IHC receptor potentials and ANF firing rates along the cochlear tonotopy and hearing range, after connecting the CoNNear cochlea [62], IHC and ANF modules together.

the fiber [90]. Experimental synchrony-level functions [50] show a non-monotonic relation to the stimulus level and exhibit maxima that occur near the steepest part of ANF rate-level curves.

Connecting the different CoNNear modules

We considered the evaluation of each CoNNear module separately, without taking into account the CoNNear models of the preceding stages and thus eliminating the contamination of the results by other factors. Each time, the evaluation stimuli were given as inputs to the reference analytical model of the auditory periphery and the necessary outputs were extracted and given as inputs to the respective CoNNear models. However, the different CoNNear models can be merged together to form different subsets of the human auditory periphery, such as CoNNear$_{\text{IHC-ANF}}$ or CoNNear$_{\text{cochlea-IHC-ANF}}$, by connecting the output of the second last layer of each model (before cropping) to the input layer of the next one. This can show how well these models can work together and how any internal noise in these neural-network architectures would affect the final response for each module. Using a CNN model of the whole auditory periphery (Fig. 11), population responses can be simulated and similar ANN-based back-ends can be added afterwards to expand the pathway and simulate higher levels of auditory processing.

References

1. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of physiology. 1952;117(4):500–544.

2. Marder E. Living Science: Theoretical musings. Elife. 2020;9:e60703.

3. Dayan P, Abbott LF. Theoretical neuroscience: computational and mathematical modeling of neural systems. MIT press; 2001.

4. Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. Journal of neurophysiology. 2001;85(3):1220–1234.

5. Miller LM, Escabi MA, Read HL, Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. Journal of neurophysiology. 2002;87(1):516–527.

6. Pozzorini C, Mensi S, Hagens O, Naud R, Koch C, Gerstner W. Automated high-throughput characterization of single neurons by means of simplified spiking models. PLoS computational biology. 2015;11(6).

7. Zeddies DG, Siegel JH. A biophysical model of an inner hair cell. The Journal of the Acoustical Society of America. 2004;116(1):426–441.

8. Lopez-Poveda EA, Eustaquio-Martín A. A biophysical model of the inner hair cell: the contribution of potassium currents to peripheral auditory compression. Journal of the Association for Research in Otolaryngology. 2006;7(3):218–235.

9. Altoè A, Pulkki V, Verhulst S. The effects of the activation of the inner-hair-cell basolateral K+ channels on auditory nerve responses. Hearing research. 2018;364:68–80.

10. Dierich M, Altoè A, Koppelmann J, Evers S, Renigunta V, Schäfer MK, et al. Optimized Tuning of Auditory Inner Hair Cells to Encode Complex Sound through Synergistic Activity of Six Independent K+ Current Entities. Cell Reports. 2020;32(1):107869. doi:10.1016/j.celrep.2020.107869.

11. Meddis R. Simulation of mechanical to neural transduction in the auditory receptor. The Journal of the Acoustical Society of America. 1986;79(3):702–711.

12. Westerman LA, Smith RL. A diffusion model of the transient response of the cochlear inner hair cell synapse. The Journal of the Acoustical Society of America. 1988;83(6):2266–2276.

13. Sumner CJ, Lopez-Poveda EA, O'Mard LP, Meddis R. A revised model of the inner-hair cell and auditory-nerve complex. The Journal of the Acoustical Society of America. 2002;111(5):2178–2188.

14. Pandarinath C, Nuyujukian P, Blabe CH, Sorice BL, Saab J, Willett FR, et al. High performance communication by people with paralysis using an intracortical brain-computer interface. Elife. 2017;6:e18554.

15. Sussillo D, Stavisky SD, Kao JC, Ryu SI, Shenoy KV. Making brain–machine interfaces robust to future neural variability. Nature communications. 2016;7:13749.

16. Klinger NV, Mittal S. Clinical efficacy of deep brain stimulation for the treatment of medically refractory epilepsy. Clinical Neurology and Neurosurgery. 2016;140:11–25.

17. Ezzyat Y, Wanda PA, Levy DF, Kadel A, Aka A, Pedisich I, et al. Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. Nature communications. 2018;9(1):1–8.

18. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436–444.

19. Kim JS, Greene MJ, Zlateski A, Lee K, Richardson M, Turaga SC, et al. Space–time wiring specificity supports direction selectivity in the retina. Nature. 2014;509(7500):331–336.

20. Yamins DL, DiCarlo JJ. Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience. 2016;19(3):356.

21. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature neuroscience. 2018;21(9):1281.

22. Steinmetz NA, Koch C, Harris KD, Carandini M. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. Current opinion in neurobiology. 2018;50:92–100.

23. Kriegeskorte N, Douglas PK. Cognitive computational neuroscience. Nature neuroscience. 2018;21(9):1148–1160.

24. Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D. Reinforcement learning, fast and slow. Trends in cognitive sciences. 2019;.

25. Kell AJ, McDermott JH. Deep neural network models of sensory systems: windows onto the role of task constraints. Current opinion in neurobiology. 2019;55:121–132.

26. Einevoll GT, Destexhe A, Diesmann M, Grün S, Jirsa V, de Kamps M, et al. The scientific case for brain simulations. Neuron. 2019;102(4):735–744.

27. Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, et al. A deep learning framework for neuroscience. Nature neuroscience. 2019;22(11):1761–1770.

28. Amsalem O, Eyal G, Rogozinski N, Gevaert M, Kumbhar P, Schürmann F, et al. An efficient analytical reduction of detailed nonlinear neuron models. Nature Communications. 2020;11(1):1–13.

29. McClelland JL, Rumelhart DE. A simulation-based tutorial system for exploring parallel distributed processing. Behavior Research Methods, Instruments, & Computers. 1988;20(2):263–275.

30. Khaligh-Razavi SM, Kriegeskorte N. Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS computational biology. 2014;10(11).

31. Kell AJ, Yamins DL, Shook EN, Norman-Haignere SV, McDermott JH. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. Neuron. 2018;98(3):630–644.

32. Bashivan P, Kar K, DiCarlo JJ. Neural population control via deep image synthesis. Science. 2019;364(6439):eaav9436.

33. Akbari H, Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N. Towards reconstructing intelligible speech from the human auditory cortex. Scientific reports. 2019;9(1):1–12.

34. Yang GR, Wang XJ. Artificial neural networks for neuroscientists: A primer. arXiv preprint arXiv:200601001. 2020;.

35. Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. Nature methods. 2018;15(10):805–815.

36. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nature neuroscience. 2009;12(5):535.

37. Gonçalves PJ, Lueckmann JM, Deistler M, Nonnenmacher M, Öcal K, Bassetto G, et al. Training deep neural density estimators to identify mechanistic models of neural dynamics. bioRxiv. 2020; p. 838383.

38. Nouvian R, Beutner D, Parsons TD, Moser T. Structure and function of the hair cell ribbon synapse. The Journal of membrane biology. 2006;209(2-3):153–165.

39. Kros C, Crawford A. Potassium currents in inner hair cells isolated from the guinea-pig cochlea. The Journal of Physiology. 1990;421(1):263–291.

40. Johnson SL. Membrane properties specialize mammalian inner hair cells for frequency or intensity encoding. Elife. 2015;4:e08177.

41. Grant L, Yi E, Glowatzki E. Two modes of release shape the postsynaptic response at the inner hair cell ribbon synapse. Journal of Neuroscience. 2010;30(12):4210–4220.

42. Chapochnikov NM, Takago H, Huang CH, Pangršič T, Khimich D, Neef J, et al. Uniquantal release through a dynamic fusion pore is a candidate mechanism of hair cell exocytosis. Neuron. 2014;83(6):1389–1403.

43. Altoè A, Pulkki V, Verhulst S. Model-based estimation of the frequency tuning of the inner-hair-cell stereocilia from neural tuning curves. The Journal of the Acoustical Society of America. 2017;141(6):4438–4451.

44. Kiang N, Baer T, Marr E, Demont D. Discharge Rates of Single Auditory-Nerve Fibers as Functions of Tone Level. The Journal of the Acoustical Society of America. 1969;46(1A):106–106.

45. Liberman MC. Auditory-nerve response from cats raised in a low-noise chamber. The Journal of the Acoustical Society of America. 1978;63(2):442–455.

46. Rhode WS, Smith PH. Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers. Hearing research. 1985;18(2):159–168.

47. Winter IM, Palmer AR. Intensity coding in low-frequency auditory-nerve fibers of the guinea pig. The Journal of the Acoustical Society of America. 1991;90(4):1958–1967.

48. Sachs MB, Abbas PJ. Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. The Journal of the Acoustical Society of America. 1974;56(6):1835–1847.

49. Relkin EM, Doucet JR. Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons. Hearing research. 1991;55(2):215–222.

50. Joris PX, Yin TC. Responses to amplitude-modulated tones in the auditory nerve of the cat. The Journal of the Acoustical Society of America. 1992;91(1):215–232.

51. Zhang X, Heinz MG, Bruce IC, Carney LH. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. The Journal of the Acoustical Society of America. 2001;109(2):648–670.

52. Heinz MG, Zhang X, Bruce IC, Carney LH. Auditory nerve model for predicting performance limits of normal and impaired listeners. Acoustics Research Letters Online. 2001;2(3):91–96.

53. Sumner CJ, Lopez-Poveda EA, O'Mard LP, Meddis R. Adaptation in a revised inner-hair cell model. The Journal of the Acoustical Society of America. 2003;113(2):893–901.

54. Meddis R. Auditory-nerve first-spike latency and auditory absolute threshold: a computer model. The Journal of the Acoustical Society of America. 2006;119(1):406–417.

55. Zilany MS, Bruce IC. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. The Journal of the Acoustical Society of America. 2006;120(3):1446–1466.

56. Zilany MS, Bruce IC, Nelson PC, Carney LH. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. The Journal of the Acoustical Society of America. 2009;126(5):2390–2412.

57. Zilany MS, Bruce IC, Carney LH. Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America. 2014;135(1):283–286.

58. Bruce IC, Erfani Y, Zilany MS. A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. Hearing research. 2018;360:40–54.

59. Verhulst S, Bharadwaj HM, Mehraei G, Shera CA, Shinn-Cunningham BG. Functional modeling of the human auditory brainstem response to broadband stimulation. The Journal of the Acoustical Society of America. 2015;138(3):1637–1659.

60. Verhulst S, Altoe A, Vasilkov V. Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. Hearing research. 2018;360:55–75.

61. Peterson AJ, Heil P. Phase locking of auditory-nerve fibers reveals stereotyped distortions and an exponential transfer function with a level-dependent slope. Journal of Neuroscience. 2019;39(21):4077–4099.

62. Baby D, Broucke AVD, Verhulst S. A convolutional neural-network model of human cochlear mechanics and filter tuning for real-time applications. arXiv preprint arXiv:200414832. 2020;.

63. Van Den Broucke A, Baby D, Verhulst S. Hearing-Impaired Bio-Inspired Cochlear Models for Real-Time Auditory Applications. Proc Interspeech 2020. 2020; p. 2842–2846.

64. Schmiedt RA. The physiology of cochlear presbycusis. In: The aging auditory system. Springer; 2010. p. 9–38.

65. Kujawa SG, Liberman MC. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. Journal of Neuroscience. 2009;29(45):14077–14085.

66. Pascual S, Bonafonte A, Serra J. SEGAN: Speech enhancement generative adversarial network. arXiv preprint arXiv:170309452. 2017;.

67. Baby D, Verhulst S. Sergan: Speech enhancement using relativistic generative adversarial networks with gradient penalty. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2019. p. 106–110.

68. Drakopoulos F, Baby D, Verhulst S. Real-time audio processing on a Raspberry Pi using deep neural networks. In: 23rd International Congress on Acoustics (ICA 2019). Deutsche Gesellschaft für Akustik; 2019. p. 2827–2834.

69. Pandey A, Wang D. Densely Connected Neural Network with Dilated Convolutions for Real-Time Speech Enhancement in The Time Domain. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2020. p. 6629–6633.

70. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–778.

71. Greenwood DD. A cochlear frequency-position function for several species—29 years later. The Journal of the Acoustical Society of America. 1990;87(6):2592–2605.

72. Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS. DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1. NASA STI/Recon technical report n. 1993;93.

73. Palmer A, Russell I. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hearing research. 1986;24(1):1–15.

74. Russell I, Cody A, Richardson G. The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown in vitro. Hearing research. 1986;22(1-3):199–216.

75. Taberner AM, Liberman MC. Response properties of single auditory nerve fibers in the mouse. Journal of neurophysiology. 2005;93(1):557–569.

76. Cheatham M, Dallos P. Response phase: a view from the inner hair cell. The Journal of the Acoustical Society of America. 1999;105(2):799–810.

77. Spoendlin H, Schrott A. Analysis of the human auditory nerve. Hearing research. 1989;43(1):25–38.

78. Jepsen ML, Ewert SD, Dau T. A computational model of human auditory signal processing and perception. The Journal of the Acoustical Society of America. 2008;124(1):422–438.

79. Osses Vecchi A, Verhulst S. Calibration and reference simulations for the auditory periphery model of Verhulst et al. 2018 version 1.2. arXiv preprint arXiv:191210026. 2019;.

80. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167. 2015;.

81. Bourien J, Tang Y, Batrel C, Huet A, Lenoir M, Ladrech S, et al. Contribution of auditory nerve fibers to compound action potential of the auditory nerve. Journal of neurophysiology. 2014;112(5):1025–1039.

82. Altoè A, Pulkki V, Verhulst S. Transmission line cochlear models: improved accuracy and efficiency. The Journal of the Acoustical Society of America. 2014;136(4):EL302–EL308.

83. Precise and Full-range Determination of Two-dimensional Equal Loudness Contours. Geneva, CH: International Organization for Standardization; 2003.

84. Rhode WS, Recio A. Study of mechanical motions in the basal region of the chinchilla cochlea. The Journal of the Acoustical Society of America. 2000;107(6):3317–3332.

85. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014;.

86. Chollet F, et al.. Keras; 2015. https://keras.io.

87. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:160304467. 2016;.

88. Ren T. Longitudinal pattern of basilar membrane vibration in the sensitive cochlea. Proceedings of the National Academy of Sciences. 2002;99(26):17101–17106.

89. Robles L, Ruggero MA. Mechanics of the mammalian cochlea. Physiological reviews. 2001;81(3):1305–1352.

90. Joris P, Schreiner C, Rees A. Neural processing of amplitude-modulated sounds. Physiological reviews. 2004;84(2):541–577.

91. Goldberg JM, Brown PB. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. Journal of neurophysiology. 1969;32(4):613–636.