

# Predicting early auditory evoked potentials using a computational model of auditory-nerve processing

**Temboury-Gutierrez, Miguel; Encina-Llamas, Gerard; Dau, Torsten**

*Published in:*
Journal of the Acoustical Society of America

*Link to article, DOI:*
10.1121/10.0025136

*Publication date:*
2024

*Document Version*
Publisher's PDF, also known as Version of record

MARCH 06 2024

# Predicting early auditory evoked potentials using a computational model of auditory-nerve processing

Miguel Temboury-Gutierrez ; Gerard Encina-Llamas ; Torsten Dau



*J. Acoust. Soc. Am.* 155, 1799–1812 (2024)
https://doi.org/10.1121/10.0025136

JASA ARTICLE

# Predicting early auditory evoked potentials using a computational model of auditory-nerve processing

Miguel Temboury-Gutierrez,[1,a] Gerard Encina-Llamas,[2,3] and Torsten Dau[1,2]

[1]*Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, DK-2800, Denmark*

[2]*Copenhagen Hearing and Balance Center, Ear, Nose and Throat (ENT) and Audiology Clinic, Rigshospitalet, Copenhagen University Hospital, Copenhagen, DK-2100, Denmark*

[3]*Faculty of Medicine, University of Vic—Central University of Catalonia (UVic-UCC), Vic, 08500, Catalonia, Spain*

**ABSTRACT:**

Non-invasive electrophysiological measures, such as auditory evoked potentials (AEPs), play a crucial role in diagnosing auditory pathology. However, the relationship between AEP morphology and cochlear degeneration remains complex and not well understood. Dau [J. Acoust. Soc. Am. **113**, 936–950 (2003)] proposed a computational framework for modeling AEPs that utilized a nonlinear auditory-nerve (AN) model followed by a linear unitary response function. While the model captured some important features of the measured AEPs, it also exhibited several discrepancies in response patterns compared to the actual measurements. In this study, an enhanced AEP modeling framework is presented, incorporating an improved AN model, and the conclusions from the original study were reevaluated. Simulation results with transient and sustained stimuli demonstrated accurate auditory brainstem responses (ABRs) and frequency-following responses (FFRs) as a function of stimulation level, although wave-V latencies remained too short, similar to the original study. When compared to physiological responses in animals, the revised model framework showed a more accurate balance between the contributions of auditory-nerve fibers (ANFs) at on- and off-frequency regions to the predicted FFRs. These findings emphasize the importance of cochlear processing in brainstem potentials. This framework may provide a valuable tool for assessing human AN models and simulating AEPs for various subtypes of peripheral pathologies, offering opportunities for research and clinical applications.

 https://doi.org/10.1121/10.0025136

(Received 27 June 2023; revised 16 February 2024; accepted 16 February 2024; published online 6 March 2024)

[Editor: Colleen G. Le Prell] Pages: 1799–1812



## I. INTRODUCTION

Auditory evoked potentials (AEPs) are recordings of synchronous neural activity in response to sound, captured using scalp electrodes in the "far field." Far-field responses have a low resolution compared to the higher spatial resolution of "near-field" responses obtained through invasive single-cell recordings in animals. To interpret these responses accurately, it is essential to understand the contributions from different neural sources excited by the stimuli. For example, in electrocochleographic (ECochG) recordings of transient stimuli like clicks or tone pulses, the compound action potential (CAP) is primarily influenced by basal auditory-nerve fibers (ANFs) with medium-to-high characteristic frequencies (CFs) (Bourien *et al.*, 2014). Another component, the summating potential (SP) observed in ECochG and auditory brainstem response (ABR) measurements, reflects contributions with different magnitudes and polarities from outer hair cells (OHCs) and inner hair cells (IHCs), as well as "spiking" and "dendritic" ANFs (Lutz *et al.*, 2022). Understanding the mechanisms underlying the generation of these electrophysiological responses is crucial, particularly in relation to auditory deficits such as hair-cell damage and neural degeneration in the auditory nerve (AN) and the auditory brainstem.

Computational models have been utilized to simulate early AEPs and explore peripheral auditory signal processing (Dau, 2003; Harte *et al.*, 2013; Rønne *et al.*, 2012; Verhulst *et al.*, 2015; Verhulst *et al.*, 2018). Dau (2003) introduced a model for generating ABR waveforms in response to transient stimuli and frequency-following responses (FFRs) to tonal stimuli. This model is based on the concept that evoked potentials recorded by remote electrodes can be theoretically represented by convolving an elementary unit waveform, or unitary response (UR) function, with the instantaneous discharge-rate function for the corresponding unit (Goldstein and Kiang, 1958). Dau (2003) employed the nonlinear computational AN model developed by Heinz *et al.* (2001) to calculate the instantaneous discharge firing rate for fibers within the frequency range from 0.1 to 10 kHz. The summed activity across CFs was convolved with a UR function derived from an averaged electrophysiological recording (in humans) at a specific pair of scalp electrodes, assuming that it reflects contributions from

[a]Email: mtegu@dtu.dk

  

different cell populations within the auditory brainstem. Simulated potential patterns were compared with experimental data under various stimulus and level conditions, including clicks, chirps designed to compensate for cochlear travel-time differences across frequency (Dau *et al.*, 2000), long-duration stimuli comprising chirps, pure tones, and slowly varying tonal sweeps. The results highlighted the significance of basilar-membrane (BM) traveling wave and AN processing effects in shaping ABRs and FFRs. Specifically, the modeling results indicated that the FFR to low-frequency (e.g., 300-Hz) pure tones primarily represents synchronized activity originating from ANFs tuned to mid and high CFs (i.e., off-CF fibers), rather than fibers tuned to frequencies around the stimulus frequency (referred to as on-CF fibers). Recent modeling studies suggested that traditional envelope-following responses (EFRs), obtained through amplitude-modulated (AM) tones, and EFRs associated with the CAP (Chen and Jennings, 2022) are likely predominantly influenced by basal (off-CF) ANFs. These fibers are distinct from those tuned to the carrier tone (Alamri and Jennings, 2023; Encina-Llamas *et al.*, 2019).

However, despite capturing various features in the ABR and FFR data, the original model proposed by Dau (2003) exhibited several discrepancies with the experimental data in terms of response patterns. For example, the model generated strong onsets in response to ramped periodic stimuli (e.g., pure tones or tonal sweeps) at high sound pressure levels (SPLs), which were much less pronounced in the corresponding experimental data. Additionally, the model failed to account for the level-dependent latency shift of ABR wave V. While the experimental data showed a latency shift of approximately 2.3 ms for a 50 dB level change, the model predicted a latency shift of less than 0.3 ms for the same level change. The predicted evoked-potential morphology strongly relies on the specific properties of the processing stages in the AN model. The model introduced by Heinz *et al.* (2001) represents an initial quantitative description of neural signal processing in the AN, inspired by the original work of Carney (1993). Over the past two decades, numerous refinements have been made to the model to provide more accurate descriptions of individual elements within the processing chain, aligning them better with physiological responses obtained in the cat AN (Bruce *et al.*, 2003; Carney *et al.*, 2015; Nelson and Carney, 2004; Zhang and Carney, 2005; Zilany *et al.*, 2009; Zilany *et al.*, 2014; Zilany and Bruce, 2006, 2007). The most recent version of the AN model was presented in Bruce *et al.* (2018).

In this study, we investigated the effects of AN processing on the formation of ABRs, FFRs, and EFRs using a computational model that predicts the far-field evoked potential patterns based on AN activity convolved with the UR function. The overall modeling approach and selected stimulus paradigms were similar to those employed in Dau (2003), with the addition of sinusoidally AM (SAM) and rectangularly AM (RAM) tones to elicit EFRs. EFRs are analogous to clinical auditory steady-state responses (ASSRs) and have been proposed to reflect IHC de-afferentation (or cochlear synaptopathy) (Shaheen *et al.*, 2015), with the large dynamic range elicited by RAM tones being particularly effective (Vasilkov *et al.*, 2021). One major change is that simulated AN responses were obtained using the model of Bruce *et al.* (2018) instead of the model of Heinz *et al.* (2001). The consequences and implications of the substantial modifications in the AN model for the formation and interpretation of the obtained AEP patterns were analyzed. Predicting early auditory responses offers a valuable tool for evaluating AN models, as these potentials represent a neural signature of peripheral neural excitation in the far field. Particularly, a computational framework capable of reproducing the key features of non-invasive evoked potentials holds the potential to enhance our understanding of the impact of different types of sensorineural impairment factors on the formation of early AEPs.

## II. METHODS

Figure 1(A) illustrates the overall structure of the model utilized in this study for predicting AEPs. The model comprises two main parts [main blocks in Fig. 1(A)]: A nonlinear computational model of the AN representing compound AN activity, and a linear UR function (independent of level and frequency) that converts the compound AN response into the AEP recorded by scalp electrodes. The AN model details are discussed below.

### A. The AN model

The AN model employed in this study was the phenomenological computational model developed by Bruce *et al.* (2018) [top block in Fig. 1(A)]. Initially based on physiological data from cats, the model was later adapted to humans to reflect properties of human frequency selectivity (Ibrahim and Bruce, 2010). The acoustic stimulus presented to the "model tympanic membrane" undergoes processing through a bandpass middle-ear transfer function. The frequency selectivity of the BM is modeled by two parallel filters serving as inputs to the signal path. The signal filter C1 represents selectivity at low and medium stimulation levels using a chirping or "time-varying narrowband" filter. This filter is adjusted by a feed-forward control path that simulates OHC function. The control wideband filter is also adjusted by the control path and is followed by an asymmetric saturating nonlinearity (NL), a low-pass (LP) filter, and a nonlinear transformation. OHC dysfunction can be introduced and controlled in this final stage (Zhang *et al.*, 2001). The second signal filter, C2, is a wideband linear filter representing BM excitation at high stimulus levels (Zilany and Bruce, 2006).

The IHC model includes a saturating NL following filter C1, which generates a direct current (dc) component in the IHC output for high-CF fibers. The stage following filter C2 (INV) shifts the phase by 180° and adjusts the relative amplitude of C1 and C2 at low, medium, and high SPLs. The two transformed signals are combined linearly and LP

https://doi.org/10.1121/10.0025136

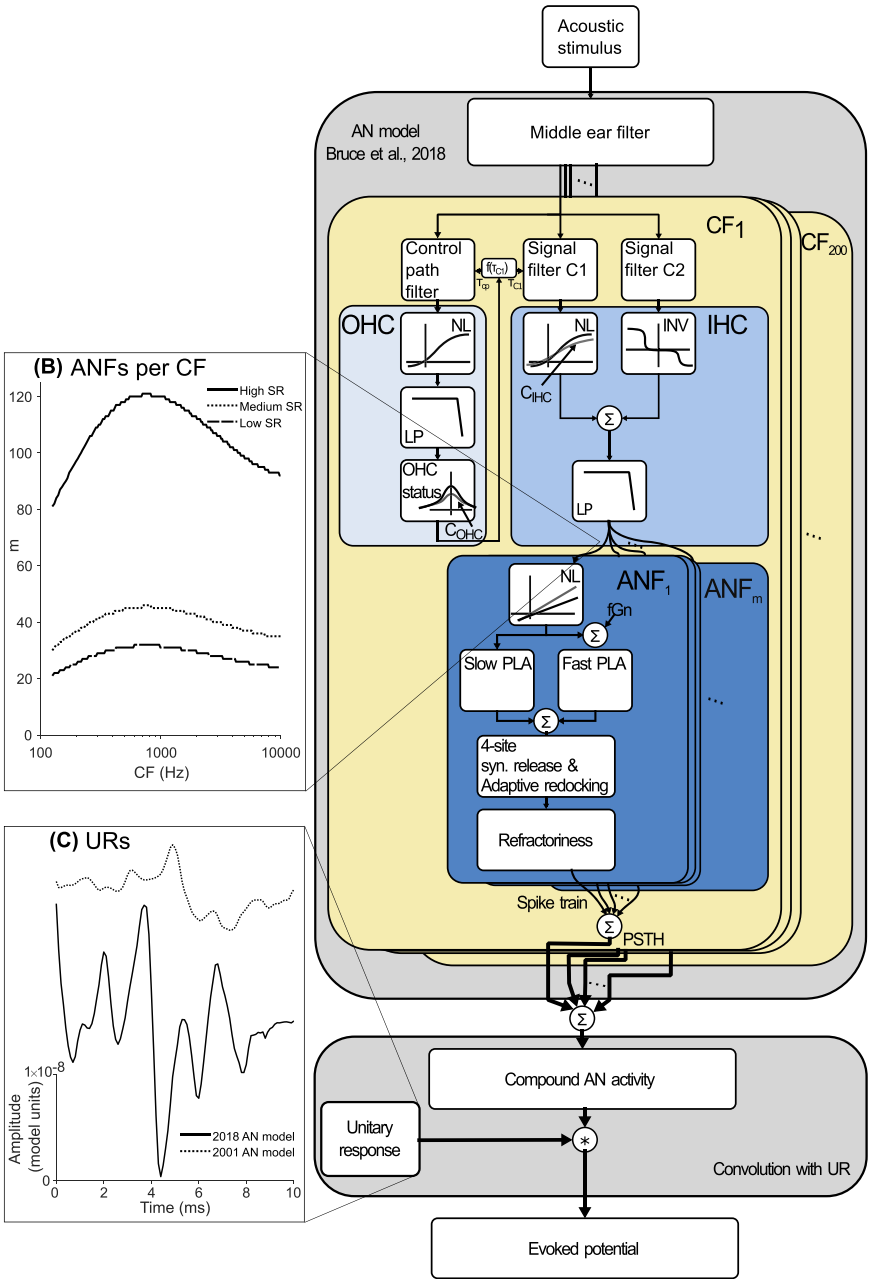


FIG. 1. (Color online) (A) Schematic of the modeling framework utilized in this study. The stimulus is first processed by the AN model of Bruce *et al.* (2018) (top block). The model includes a middle-ear filter and a control path that represents the active processes, depicted through OHC transduction. It also incorporates a signal path representing IHC transduction and an ANF model that includes synaptic transmission and a spike generator. The AN model is implemented by simulating 200 IHCs, with each one being assigned a unique CF. Each IHC is innervated by a varying number of ANFs, represented by "m," in accordance with physiological data from humans as documented by Spoendlin and Schrott (1989). The spike trains across the ANFs are summed, and the compound AN response is convolved (represented by the symbol *) with a unitary response (UR). This process linearly transforms spikes into the AEP recorded in the far field (as depicted in the bottom block). Key terms include the following: Time constant (T), low-pass (LP) filter, static nonlinearity (NL), characteristic frequency (CF), inverting nonlinearity (INV), auditory nerve fiber (ANF), power law adaptation (PLA), fractional Gaussian noise (fGn), and post-stimulus time histogram (PSTH). $C_{OHC}$ and $C_{IHC}$ are scaling constants that specify OHC status and IHC status, respectively. (B) Number of ANFs (m) per CF, divided into high-, medium-, and low-spontaneous rate (SR) fibers, based on data from Spoendlin and Schrott (1989). (C) URs derived for the 2018 and 2001 AN model simulations.

filtered to limit sensitivity to phase locking (Carney, 1993) and represent the relative IHC transmembrane potential.

The synaptic transmission between IHCs and ANFs and the spike generation of the ANFs are modeled using a "gently saturating" nonlinear mapping function, designed to achieve realistic ANF thresholds and dynamic ranges in the AN rate-level functions with respect to CF and spontaneous rate (SR) (Bruce *et al.*, 2018). This stage is followed by a slow power-law adaptation (PLA) function representing offset adaptation and a fast PLA function simulating the effects of "additivity" in AN rate responses to stimuli with amplitude increments (Zilany *et al.*, 2009). Finally, absolute and relative refractory periods of the ANF are incorporated. The model generates a spike train for each CF, simulating the response of a single ANF tuned to that CF.

The AN model was implemented following the description in studies by Encina-Llamas *et al.* (Encina-Llamas *et al.*, 2019; Encina-Llamas *et al.*, 2021). In all AN simulations presented in this study, 200 CFs were utilized, ranging from 125 Hz to 10 kHz and distributed uniformly by distance across the BM based on the function of Greenwood (1990). A single IHC was simulated for each CF, resulting in a total of 200 simulated IHCs. For each CF, multiple ANFs were simulated, following a non-uniform distribution estimated from the healthy human AN (Spoendlin and Schrott, 1989) [Fig. 1(B)], resulting in 32 000 ANFs. Consequently, in whole-nerve simulations, multiple ANFs tuned to the same CF were simulated. The spike trains of each ANF per CF were summed to obtain a population response, similar to a post-stimulus time histogram (PSTH),

for each CF. Finally, population responses from each CF were summed across all CFs to derive a compound AN response. The model of Bruce *et al*. (2018) allows for the input of SR as a free parameter, enabling the simulation of different SR responses. The distribution of high-, medium-, and low-SR fibers was set at 61%, 23%, and 16%, respectively, based on data from cat ANFs (Liberman, 1978), as human data are not available. The AN model operated at a sampling rate of 100 kHz.

### B. Main changes between the 2018 model of Bruce *et al.* and 2001 model of Heinz *et al.*

Dau (2003) employed the AN model by Heinz *et al.* (2001) as the front-end of the modeling framework for predicting AEPs. Over the past 20 years, several modifications have been made to the AN model to enhance the accuracy of the physiological response simulations. Notably, the implementation of PLA has been a significant change, describing neural offset and recovery adaptation and long-term dynamics in the IHC-synapse complex. Additionally, PLA accounts for the effect of "additive" rates to amplitude increments, which reflects the ANF firing rates in response to level increases or decreases, independent of the temporal interval between onset and level change. This effect enhances response synchrony to low-frequency tones and AMs, limited by the IHC cutoff frequency (Zilany *et al*., 2009). Bruce *et al*. (2018) incorporated a new synapse model utilizing four synaptic vesicle-docking sites, based on recent physiological data (Peterson *et al*., 2014; Peterson and Heil, 2018). This improvement enhanced the simulation of the ANF spiking statistics, particularly in physiological forward masking, resulting in more realistic AN rate-level functions. Furthermore, the model accounted for rapid (2-ms) and short-term (60-ms) exponential adaptation observed in onset responses. Other major modifications included more accurate responses to high-intensity stimuli (Zilany and Bruce, 2006), enabling the model to account for highly nonlinear AN response properties such as peak splitting (Kiang, 1990). Redesign of the IHC transduction and the IHC-AN synaptic complex (Zhang and Carney, 2005) improved maximum rates at saturation for high-CF fibers (Zilany *et al*., 2014), and more precise effects of acoustic trauma (Bruce *et al*., 2003) were also implemented. The AN model was also adjusted to represent human characteristics (Ibrahim and Bruce, 2010) by modifying certain model free parameters. These modifications included the middle-ear filter transfer function, ANF frequency tuning based on psychoacoustic estimates (Glasberg and Moore, 1990) and human otoacoustic emissions (Shera *et al*., 2001), and the audible frequency range.

### C. The UR

The transformation from compound AN activity to far-field evoked potentials is modeled by convolving the population AN response with a UR function (Dau, 2003; Melcher and Kiang, 1996). This approach is based on the original work of Goldstein and Kiang (1958), where the linear transformation from the spontaneous discharge of a given neuron to its corresponding activity at electrodes located in the far field was derived through deconvolution. The response captured by the electrodes was repeatedly deconvolved with the neuron's discharge pattern, effectively averaging out neural noise that was not related to the neuron's activity. By convolving the neuron's response pattern with the UR function, activity measured from these electrodes, reflecting the neuron's response to acoustic stimulation, could be predicted. This concept was extended from individual cells to groups of cells based on established anatomical and physiological knowledge (Melcher and Kiang, 1996). Therefore, instantaneous discharge rates from the entire AN in response to acoustic stimulation are linearly transformed using a single $UR_{AN}$ to simulate activity in the far field corresponding to AN activity. To obtain the brainstem's response, such as the ABR, to a given stimulus, only the UR function and the compound discharge pattern of each neural generator are necessary. Dau (2003) further expanded this concept and proposed using a single UR function to transform the compound AN activity pattern into a full ABR waveform containing waves generated by multiple nuclei. Thus, this UR represents not only the propagation of the electric field generated by the compound AN activity to the recording site, similar to $UR_{AN}$, but also includes contributions from subsequent nuclei. The main argument was that spherical bushy cells in the anterior ventral cochlear nucleus (CN) and principal cells in the medial superior olive (MSO), believed to be the primary generators of ABR waves III and V, respectively (Scherg and Von Cramon, 1985), exhibit firing behavior similar to that of primary-like neurons (Blackburn and Sachs, 1989; Goldberg, 1975; Yin and Chan, 1990). In Dau (2003), the UR function was obtained by deconvolving averaged ABRs obtained experimentally from several normal-hearing listeners with the simulated compound AN response (summed across CFs) using the AN model. The UR function used in the present study was derived using the same experimental ABR at 96 dB peak-equivalent sound pressure level (peSPL) from Dau (2003). In future work, the UR could be derived using ABRs calculated with a different number of averages, allowing the measurement noise to be parametrized in the selected UR. In this study, the final simulated waveforms are presented without [black, thick traces in Figs. 4(B), 4(C), 5, and 6] and with added internal noise, similar to Dau (2003) [gray, thin traces in Figs. 4(B), 4(C), 5, and 6). White Gaussian noise was chosen as internal noise, and the power was adjusted to correspond to a signal-to-noise ratio (SNR) of 5 dB at the highest stimulation level. The URs derived in this study for the 2018 and 2001 AN model simulations are shown in Fig. 1(C). Note that the UR for the 2018 model is larger, since the output of the 2001 model to the click is larger.

### D. Stimulus conditions

The stimulus conditions were similar to those considered in Dau (2003). Data were taken from Dau *et al*. (2000). Transient-evoked responses, specifically ABRs, were

https://doi.org/10.1121/10.0025136

simulated using 80-$\mu$s clicks presented at stimulus levels ranging from 46 to 96 peSPL in 10 dB increments. Responses to sustained stimuli, known as FFRs, were simulated using 50-ms, 300-Hz pure tones with 5-ms Hanning ramps at the onset and offset. The tones were presented at SPLs of 50 to 100 dB in 10 dB steps. Moreover, the "combination stimulus" from Dau *et al.* (2000) was considered, comprising two 30-ms pure tones at 320 Hz and 8 kHz and a temporally and spectrally embedded chirp with an instantaneous frequency rising from 320 to 8000 Hz. This chirp was designed to compensate for BM travel-time differences across frequencies. The chirp was embedded in the waveform between the low-frequency tone and the chirp onset as well as the chirp offset and the high-frequency tone. This stimulus was included to investigate potential distinctions in the modeling framework's response to onset transients as opposed to transients embedded in an acoustic stimulus. In Dau (2003), the framework produced exaggerated onset responses, even to ramped stimuli; however, the simulated response to the embedded chirp was qualitatively more accurate. Finally, 4000 Hz pure tones with an amplitude modulated at 120 Hz, either with a sinusoidal (SAM tones) or a rectangular (duty cycle = 25% [RAM tones]) modulator at 85 dB rmsSPL, were used to simulate EFRs, as in Vasilkov *et al.* (2021). All stimuli were generated digitally in MATLAB (The MathWorks) at a sampling frequency of 48 kHz. The stimulus waveforms were resampled to match the model sampling frequency (100 kHz) with the MATLAB function *resample()*.

## III. RESULTS

### A. Simulations of single-unit AN (near-field) responses to tones and clicks in cats

The AN models were originally designed to simulate AN responses in cats. Figure 2 illustrates simulations of AN responses to pure tones and clicks in cats using two different models: the AN model of Bruce *et al.* (2018), referred to as the "2018 AN model" (depicted by red circles), and the model of Heinz *et al.* (2001), referred to as the "2001 AN model" (depicted by blue triangles). These simulations are compared with physiological recordings from Kim *et al.* (1990) and Lin and Guinan (2000) (depicted by black squares).

Figures 2(A) and 2(C) display responses to tones at various SPLs. The tones used were 200-ms 1-kHz pulses with 5-ms onset and offset ramps, similar to the study conducted by Kim *et al.* (1990), at levels of 30, 50, and 70 dB SPL. The driven rate, which represents the average firing rate minus the average SR, is shown as a function of the CF of high-SR ANFs. At 30 dB SPL [top panel in Fig. 2(C)], the recorded data (squares) demonstrated a frequency-specific neural response centered around the stimulus frequency (1 kHz), with a maximum driven-rate response of approximately 180 spikes/s. As the SPL increased, the AN response spread mainly towards higher CFs, indicating an "upward spread" of neural excitation. The simulated responses obtained with both AN models captured the on-CF response observed in the data and reasonably reflected the neural response at CFs below the stimulation frequency. However, deviations from the data were observed for CFs above the stimulus frequency in the simulations. Specifically, the 2001 AN model exhibited a substantially larger neural response compared to the data, particularly at the highest stimulus level (70 dB). In contrast, the simulations obtained with the 2018 AN model provided a closer match to the data. The experimental and simulated data were fitted using the *smooth* function in MATLAB, employing local regression with weighted linear least squares and a second-degree polynomial model. Outliers in the regression were assigned lower weights. The fitted simulations were evaluated computing the Euclidean distance using the *pdist2* MATLAB function.

Figures 2(B) and 2(D) display measured and simulated responses to 90-$\mu$s clicks presented at levels ranging from 64 and 114 dB peSPL, in 5 dB increments, similar to the study of Lin and Guinan (2000). Clicks of both polarities (condensation and rarefaction) were presented, and the PSTH was obtained after 500 repetitions at each level. The earliest peak latency for each CF among the responses for all presentation levels and polarities was selected, both in the recordings (Lin and Guinan, 2000) and in the simulations. Similar to Fig. 2(A), the data were smoothened using local regression with weighted linear least squares (solid lines). The simulations conducted with the 2018 AN model (circles) exhibit a very close correspondence to the experimental data (squares), with a small deviation observed at the highest frequencies (>10 kHz). In contrast, the simulations performed with the 2001 AN model (triangles) indicate very small latency values around 0.5 ms, demonstrating minimal variability across CF compared to the values between 1 and 3.5 ms observed in the data as well as in the simulations with the 2018 AN model. This nearly flat latency-CF function obtained with the 2001 AN model suggests that the temporal response at the output of this model to transient broadband stimuli such as clicks or to onsets of sustained stimuli like tones overestimates the amount of neural synchronization across CFs.

### B. Simulated far-field evoked potentials in humans

Figure 3 shows the experimental and simulated ABRs to clicks at various levels of stimulations. Figure 3(A) displays the recorded responses obtained from Dau (2003), with the cross indicating the peak of wave V for each ABR. The dashed line represents the latency-level function derived from the observed wave-V latencies. Figures 3(B) and 3(C) present the corresponding simulated responses using the 2018 and 2001 AN models, respectively, integrated into the overall evoked-potential prediction framework, as depicted in Fig. 1. For the simulations, the UR functions were derived from the experimental ABRs obtained in response to the click at the highest level (96 dB peSPL). Since this was the fitting condition, it was expected

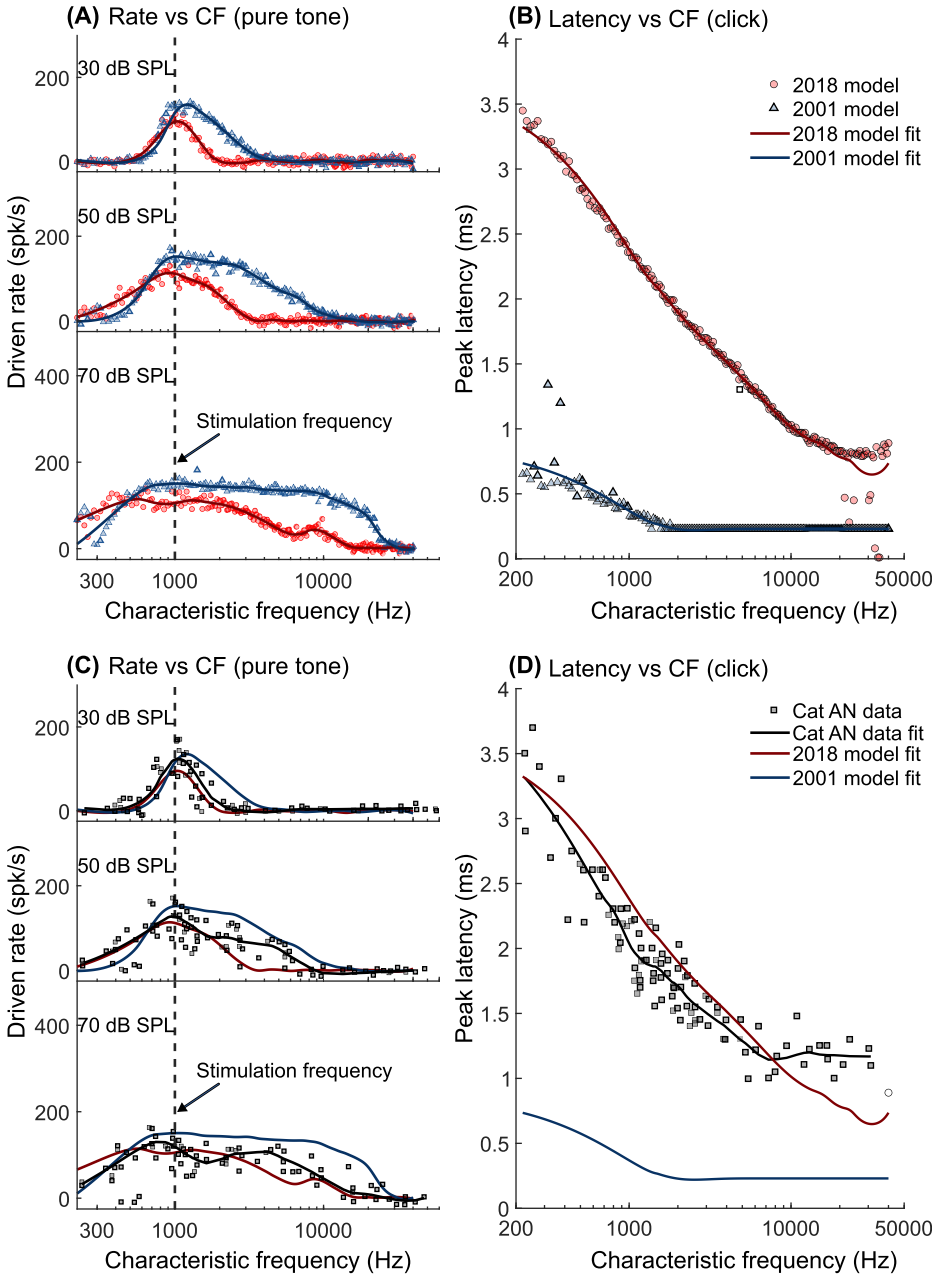


FIG. 2. (Color online) Measured and simulated responses of high-SR ANF responses to pure tones (A and C) and clicks (B and D) in cats. (A) Simulated ANF driven rates (2018 AN model, circles; 2001 AN model, triangles) in response to 200-ms, 1-kHz pure tones at 30, 50, and 70 dB SPL (top, middle and bottom, respectively) as a function of CF. The driven rates were computed by subtracting the SR during the 800- to 1000-ms period from the firing rates during the 0- to 200-ms window. (B) Simulated ANF earliest peak latencies (2018 AN model, circles; 2001 AN model, triangles) in response to 90-$\mu$s clicks at 64–114 dB peSPL, presented in 5 dB steps, using 500 repetitions per level with both polarities at 33.3 Hz. The solid lines represent smoothed fitted functions obtained through local regression with weighted linear least squares and a 2nd degree polynomial (2018 model, red; 2001 model, blue). (C) Experimental data (Kim *et al.*, 1990) (squares) and model fits to same stimuli as in panel (A). (D) Experimental data (Lin and Guinan, 2000) (squares) and model fits to same stimuli as in panel (B).

that at this level the experimental and simulated ABRs were similar. At lower levels, the simulations generated by both models exhibited ABR morphologies that also closely resembled those observed in the experimental data. While both models predicted ABR wave-V peak amplitudes similar to those observed in the measured data, the simulations obtained with the 2001 AN model demonstrated a slower decay of wave-V amplitude with decreasing level. This resulted in a distinct peak at the lowest simulated level at 46 dB peSPL, which was not present in the data or in the predictions obtained with the 2018 AN model. Regarding wave-V latency, the data showed an increase in 1.84 ms for a 40 dB level change, consistent with findings from other studies (Jiang *et al.*, 1991; Picton *et al.*, 1981; Prosser and Arslan, 1987; Serpanos *et al.*, 1997; Strelcyk *et al.*, 2009). However, both models strongly underestimated the observed latency change and only predicted an effect of approximately 0.4 ms within the considered level range.

Figure 4 shows the measured and simulated responses to the "combination stimulus." The top row of the figure displays the waveform of the stimulus, with vertical lines indicating the onset and offset of the embedded chirp. Figure 4(A) presents the data obtained at different stimulation levels, which have been replotted from Dau (2003). It is evident that there were phase-locked responses to the low-frequency tone (320 Hz) in the first segment of the stimulus. Additionally, a response to the embedded chirp can be observed, particularly at lower stimulation levels. The chirp elicited a synchronized neural response with a latency relative to the chirp offset, corresponding to a wave-V

https://doi.org/10.1121/10.0025136

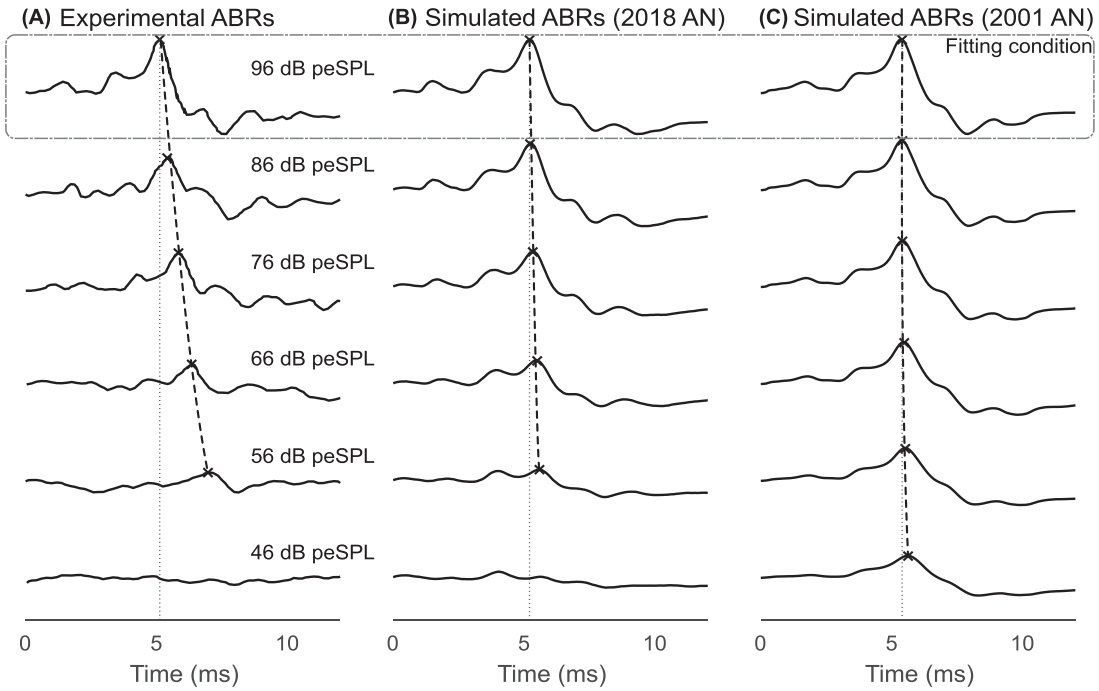


FIG. 3. Measured and simulated click-evoked auditory brainstem responses (ABR) in human listeners. (A) Recorded ABR data from Fig. 2 in Dau *et al.* (2000) in response to clicks presented at 46, 56, 66, 76, 87, and 96 dB peSPL. (B) Simulated ABRs to the same stimuli using a UR derived via deconvolution from the recorded ABR obtained at the highest stimulation level and the 2018 AN model. (C) Corresponding simulations obtained with the 2001 AN model. Wave-V peaks are indicated by crosses, and the latency-level functions are represented by dashed lines. Dotted vertical lines have been added as a visual aid, indicating the latency of the wave-V peaks at the 96-dB-peSPL level. The dashed-dotted rectangle indicates the condition in which the UR was derived (i.e., the fitting condition).

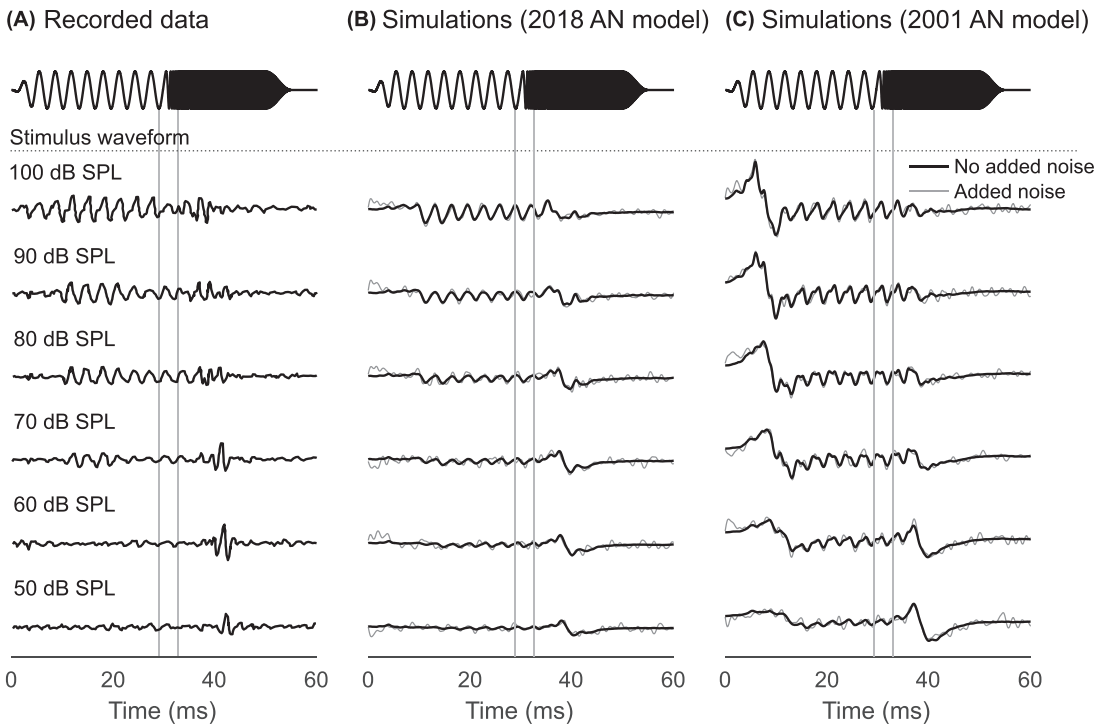


FIG. 4. Measured and simulated evoked responses to the "combination stimulus," which consists of a low-frequency (300-Hz) tone segment, followed by a rising chirp (between 0.3 and 8 kHz) designed to compensate for BM travel-time differences across frequency, and finally, a high-frequency (8-kHz) tone segment, as indicated in the top row of each panel. The onset and offset of the embedded chirp are indicated by vertical lines. (A) Recorded responses, replotted from Dau (2003), to the stimulus presented at 50, 60, 70, 80, 90, and 100 dB SPL. (B) Simulated responses to the combination stimulus using the 2018 AN model. Thick black lines represent simulations without added noise. Thin gray lines illustrate simulations with added Gaussian noise representing measurement noise. (C) Corresponding simulations obtained with the 2001 AN model.

component, as described in previous studies (Dau *et al.*, 2000; Fobel and Dau, 2004; Wegner and Dau, 2002).

Panels (B) and (C) of Fig. 4 show the corresponding simulated responses obtained with the 2018 and 2001 AN models, respectively. Consistent with the data, both models exhibited strong phase-locked responses to the 320-Hz tone at high stimulation levels and no FFR to the high-frequency tone (8 kHz) in the final segment of the stimulus. However, the 2001 AN model [Fig. 4(C)] produced exaggerated onset responses that were absent in the data as well as in the simulations obtained with the 2018 AN model. Regarding the embedded chirp, both models qualitatively accounted for the observed 'transient' response in the data at lower stimulation levels. However, similar to the results described in Fig. 3, both models again clearly underestimated the latency change with stimulation level.

To investigate the generation mechanisms of the FFR to pure tones within the model framework, Dau (2003) separated the simulated evoked potentials into responses resulting from contributions of low- and high-frequency fibers in the AN. Following a similar approach, Fig. 5 illustrates the simulated FFRs to a 50-ms, 300-Hz tone using the 2018 AN model [Fig. 5(A)] and the 2001 AN model [Fig. 5(B)]. To demonstrate the contributions of "low-frequency" versus "high-frequency" AN activity to the predicted response within the model, the summed responses from the 3-octave wide low-frequency region (0.125 to 1 kHz; left) and the 3-octave wide high-frequency region (1 to 8 kHz; right) are shown. Similar to Dau (2003), at SPLs below 70 dB peSPL, simulated FFRs with added noise (thin gray traces) were not distinguishable, in line with experimental observations. Despite this similarity, the results obtained with the two model versions exhibited substantial differences. The 2018 AN model [Fig. 5(A)] predicted prominent "phase-locked" responses in both frequency regions at the highest SPLs. As SPLs decrease, the simulated AN FFR became more pronounced in the lower-frequency region compared to the higher-frequency region. On the other hand, the 2001 AN model [Fig. 5(B)], as discussed in Dau (2003), exhibited a response that was clearly dominated by high-CF activity at all SPLs above 70 dB. In this model, the low-frequency region did not produce any stimulus-related periodicity in the response, even at the highest SPLs. Additionally, as observed in Fig. 4(C), the 2001 AN model generated exaggerated onset responses in both frequency regions, which were not observed in the simulations obtained with the 2018 AN model.

To further investigate the differences between the simulations shown in Fig. 5 obtained with the two models, AN responses at different CFs (indicated in hertz) and their contribution to the far-field compound response were analyzed. Figure 6 presents simulated FFRs to a high-intensity 100-dB peSPL 300-Hz tone. Figure 6(A) shows the results obtained with the 2018 AN model, while Fig. 6(B) shows the

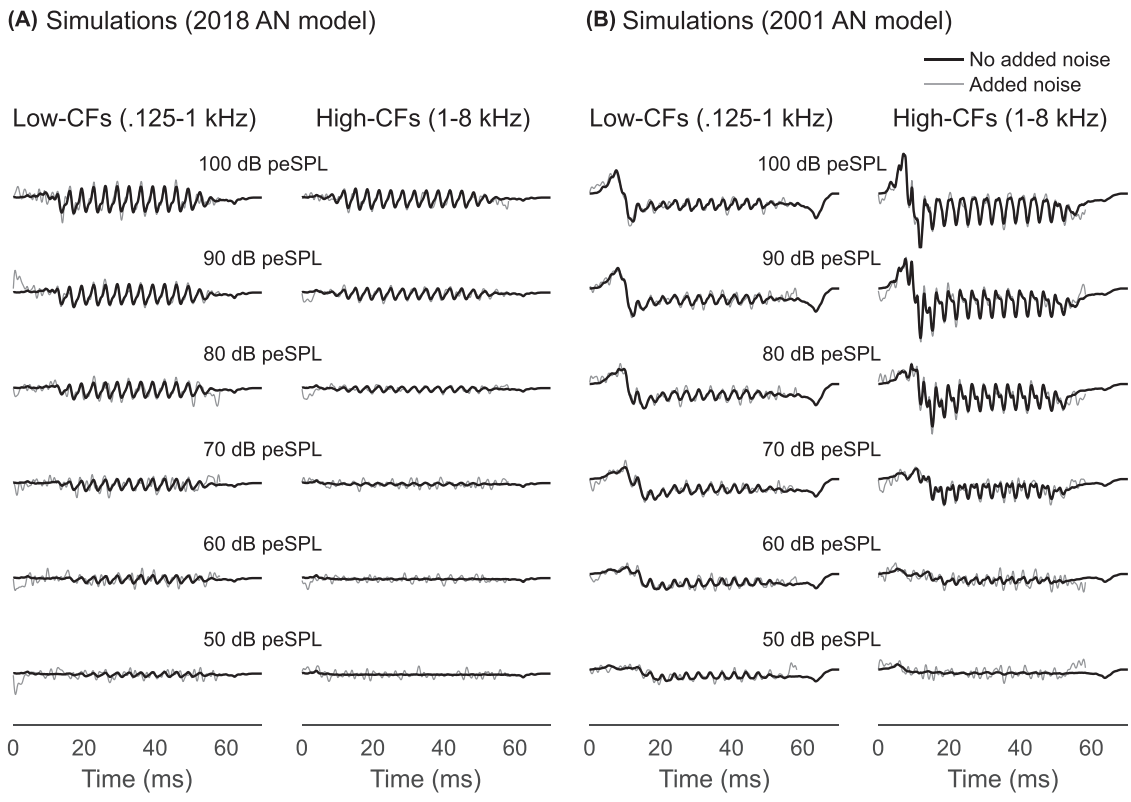


FIG. 5. Simulated responses to a 300-Hz tone at 50, 60, 70, 80, 90, and 100 dB SPL obtained with the 2018 AN model (A) and the 2001 AN model (B). (A) Contributions from low-frequency (0.125–1 kHz, left) vs high-frequency (1–8 kHz, right) AN activity to the tone at different SPLs, obtained using the 2018 AN model. Thick black lines represent simulations without added noise. Thin gray lines illustrate simulations with added Gaussian noise representing measurement noise. (B) Corresponding simulations obtained with the 2001 AN model.

https://doi.org/10.1121/10.0025136

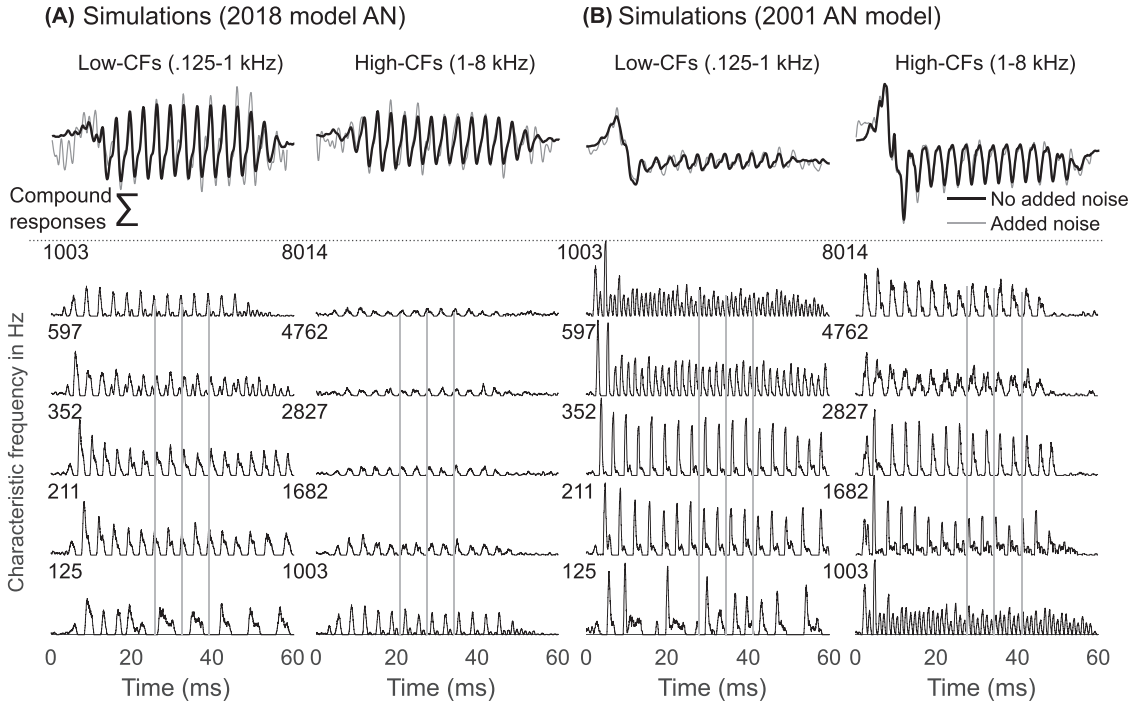


FIG. 6. Simulated responses to a 300-Hz pure tone at 100 dB peSPL obtained with the 2018 AN model (A) and the 2001 AN model (B). Contributions from low-frequency (0.125–1 kHz; left) vs high-frequency (1–8 kHz; right) AN activity are shown in both panels. (A) The top rows display the far-field response of all the ANFs in the two frequency ranges using the 2018 AN model. Thick black lines represent simulations without added noise. Thin gray lines illustrate simulations with added Gaussian noise representing measurement noise. The bottom rows show the response of single CFs (representing a group of ANFs) at different frequencies (125, 211, 352, 597, 1003, 1682, 2827, 4762, and 8014 Hz, with the same scale for both models). (B) Corresponding simulations obtained with the 2001 AN model.

corresponding results obtained with the 2001 AN model. Similar to Fig. 5, the left and right columns in the two panels represent the contributions in the lower-frequency region (between 0.1 and 1 kHz) and the higher-frequency region (between 1 and 8 kHz), respectively. In the top section of each panel, the FFRs obtained by convolving the responses summed across CFs for the two frequency regions are displayed. For the 2018 AN model [Fig. 6(A)], the compound neural activity (top) resulting from the low-frequency region was comparable to that obtained from the higher-frequency region. The single-CF responses (bottom rows) exhibited relatively strong activity at low frequencies around the stimulus frequency (300 Hz), but lacked strong synchronization across frequency due to travel-time differences on the BM. At the higher frequencies, the single-channel response amplitudes were considerably smaller but demonstrated a higher degree of synchrony across channels. Consequently, the compound response for this stimulus at a high stimulus level showed a similar overall amplitude. Therefore, according to this model, the FFR to a low-frequency high-intensity tone is generated by numerous ANFs tuned to a wide range of CFs. In contrast, the compound responses obtained with the 2001 AN model [Fig. 6(B)] suggested that the FFR is primarily generated by the synchronized neural activity in the higher-frequency channels, as discussed in Dau (2003). Furthermore, the responses obtained with the 2001 AN model exhibited "oscillations" at higher frequencies (not matching the 300-Hz stimulus frequency) that were also present in the simulated compound response. These oscillations were not observed in the data nor in the predictions with the 2018 AN model. Last, as described earlier (Figs. 3 and 4), the 2001 model predicts stronger onset responses than those observed in the FFR data and in the 2018 model.

EFRs have been proposed to be a marker for IHC deafferentation and are similar to the clinical ASSR (Shaheen *et al.*, 2015; Keshishzadeh *et al.*, 2020). RAM tones have been shown to produce larger responses than SAM tones (Vasilkov *et al.*, 2021) [Fig. 7(C), in gray]. The time-domain waveform (left) and spectrum (right) of the SAM and RAM tones used in this study are shown in Figs. 7(A) and 7(B), respectively. The simulated EFRs, modeled using the 2018 AN model, are shown in dark blue. The contributions from on-CF ($\frac{1}{3}$ octave around 4 kHz) and basal (5–10 kHz) and more apical (1–2 kHz) off-CF channels are shown below the EFR derived with all frequencies. Simulated RAM tones elicited larger responses than the SAM tone, as observed in the time-domain [Fig. 7(A) vs Fig. 7(B)] and larger spectral amplitudes at the modulation frequency (120 Hz) and its first 4 harmonics (up to 600 Hz, as in Vasilkov *et al.* (2021) [Fig. 7(C), in blue]. The more complex harmonic content in the stimulus of the RAM tone resulted in less sinusoidal responses compared to responses to the SAM tone [Figs. 7(B) and 7(D)], which qualitatively matched experimental EFRs to a RAM tone from a representative subject (Vasilkov *et al.*, 2021) [Fig. 7(D)].

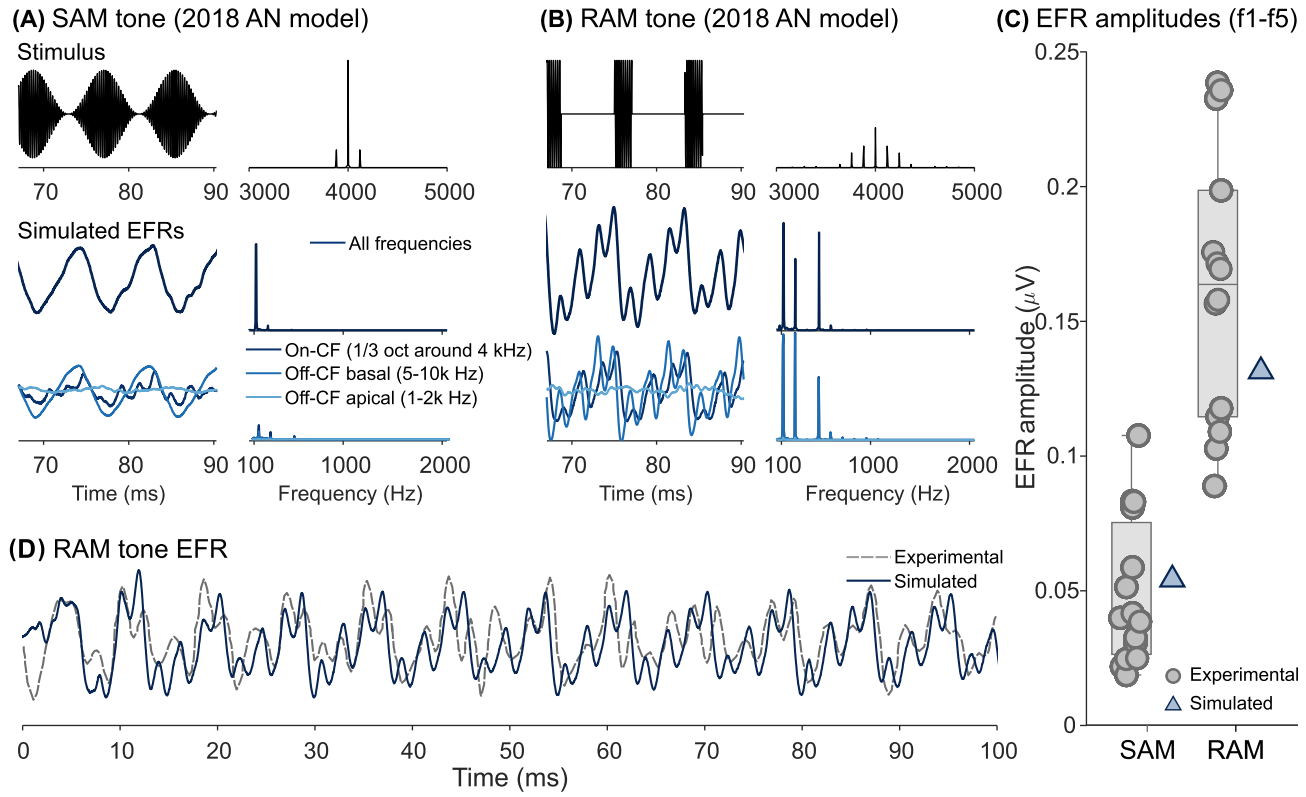


FIG. 7. (Color online) Simulated responses were generated using the 2018 AN model for a 4-kHz tone at 85 dB rmsSPL modulated at 120 Hz with a sinusoidal and rectangular (duty cycle = 25%) modulators at 100% modulation depth. (A) Time-domain (left) and frequency-domain (right) representations of the SAM tone stimulus and simulated EFRs. The modeled EFR is shown in dark blue, with the contribution of on-CF channels ($\frac{1}{3}$ octave centered around 4 kHz), off-CF basal channels (5–10 kHz) and off-CF apical channels (1–2 kHz) represented in various shades of blue below. The contribution from sets of CFs was normalized to account for the difference in bandwidths. (B) Corresponding stimulus and simulated EFRs to the RAM tone in blue. (C) Experimental EFR amplitudes (sum of f1–f5) to SAM and RAM stimuli (in gray) (Vasilkov *et al.*, 2021) and amplitudes of simulated EFRs to SAM and RAM tones in blue. (D) Temporal response to a RAM tone of a representative subject (Vasilkov *et al.*, 2021) superimposed on the simulated RAM EFR.

## IV. DISCUSSION

This study presents a model to predict early AEPs in response to various stimuli, encompassing both transient stimuli (such as clicks and chirps) and sustained stimuli (including pure tone and AM tone), across different levels of stimulation. The model builds upon a previous approach presented in Dau (2003), but incorporates a significantly modified AN processing front-end as part of the AEP prediction framework. The predicted patterns of electrical potentials obtained with the improved AN model (Bruce *et al.*, 2018) exhibited various characteristics consistent with experimental data. The specific properties of AN processing, which are highly nonlinear, strongly influence the formation of the simulated AEPs, especially at high sound levels where neural excitation is broad. Notably, the predictions obtained with the 2018 AN model demonstrated realistic onset responses to tones, while the 2001 AN model exhibited exaggerated onset responses that increased with stimulation level (Figs. 4, 5, and 6). This exaggerated response was attributed to nearly constant first-peak latencies of AN fiber in response to clicks at high levels, which were less dependent on CF [as depicted in Fig. 2(B)]. The reduced "CF-dependency," particularly at higher CFs, resulted in large compound responses arising from highly synchronized broadband AN activation, particularly in response to high-intensity onsets. In contrast, the 2018 AN model incorporated cochlear delays that exhibited a more pronounced CF-dependency both at high [Figs. 2(B) and 2(D)] and lower (not shown) levels, which aligns more closely with physiological data. Regarding click stimulation, the predictions obtained with the 2001 AN model did not exhibit an "exaggerated" onset since this model was calibrated to accurately reproduce click-ABR waveform morphologies when convolved with the UR.

One noticeable discrepancy between the data and the predictions obtained with both AN models was observed in the latency-level functions of the click-evoked (or chirp-evoked) wave V (also noted in Dau (2003). Both models clearly underestimated the change in wave-V latency as a function of SPL (see Fig. 3 for clicks and Fig. 4 for the embedded chirp). This discrepancy could potentially be attributed to retro-cochlear frequency- and level-dependent neural delays, which are not accounted for in this simplified model that assumes constant delays between wave I (generated in the AN), wave III, and wave V (Dau, 2003). One possible way to further evaluate this aspect would be to incorporate more detailed neural processes, representing the firing rates at the assumed neural generation sites at the CN,

MSO, and inferior colliculus levels, as well as UR components that reflect the transformation from the respective generators to the sensors (scalp electrodes). A recent study employed a different approach to assess the AN model, focusing on simulating cochlear-generated AEPs measured with electrocochleographically (Alamri and Jennings, 2023). Using a UR function solely reflecting AN activity, this method allowed for evaluating the model's ability to predict wave-I latency as a function of level. Latency-level functions of click-evoked CAP $N_1$ simulated with the AN models of Zilany *et al.* (2014) and Verhulst *et al.* (2018) were similarly underpredicted. The authors proposed that the level-dependent human cochlear tuning parameters, estimated in both cases with otoacoustic emissions, are not fully captured in the model (Alamri and Jennings, 2023; Verhulst *et al.*, 2015). This suggests that the latency-level mismatch lies in the process of humanizing the original cat AN model (Ibrahim and Bruce, 2010). It is plausible that the latency-level functions are physiologically accurate in cats, but the transition from cat to human may not be accurately represented. Indeed, the humanized AN model has not been validated against single-unit physiological AN data for humans due to ethical constraints. Intraoperative recordings of AN activity in humans during surgery could potentially overcome these limitations (Huet *et al.*, 2022).

Another significant disparity between the simulated potentials obtained with the two preprocessing models was the relative contributions of low-frequency versus high-frequency AN activity to the predicted FFRs (Fig. 5). The 2001 AN model generated synchronized neural activity primarily originating from high CFs (above approximately 1 kHz), even when stimulated with low-frequency tones (at 300 Hz). This indicated that FFRs are primarily influenced by basal "off-CF" neural excitation, with lower-frequency activity having limited effectiveness in contributing to FFRs. This limited contribution was due to desynchronized neural activity across the lower-frequency channels caused by larger travel-time differences at those frequencies (Dau, 2003). In contrast, the 2018 model exhibited stronger responses in the lower-frequency region, resulting in larger compound responses across frequencies in that range, especially at low and medium stimulation levels, despite the differences in energy distribution across frequencies. At high stimulus levels, the contribution from high-frequency ("off-CF") activity increased and led to an overall "broadband response" (e.g., at 100 dB peSPL) (Fig. 6). This contrasts with the interpretation provided in Dau (2003) based on the original model predictions. The simulated results of the 2018 model align with physiological data from cats, showing similar rates to pure tones at higher levels (Kim *et al.*, 1990) [Fig. 2(C); 70 dB SPL] and comparable synchrony to the carrier frequency of SAM tones in both on-CF and off-CF ANFs (Joris and Yin, 1992). The substantial differences between the predictions obtained with the two models primarily stem from several factors: the re-mapping of the response magnitude as a function of CF (Zilany *et al.*, 2014), the improved CF- and SR-dependent rate-level functions (Bruce *et al.*, 2018), and the elimination of a static nonlinearity in the control path (Bruce *et al.*, 2003). The presence of a static compressive nonlinearity in the 2001 AN model [between the control path filter and OHC nonlinearity, as shown in Zhang *et al.* (2001)] (Fig. 1), caused undesired distortion products or "oscillations" in the AN response (Bruce *et al.*, 2003). These oscillations, observed in responses at low-to-mid CFs (~500–1000 Hz) in the 2001 AN model (Fig. 6), resulted in weaker phase-locked compound potentials (due to destructive interference) compared to the responses originating from higher CFs. In contrast, in the 2018 AN model, the low-to-mid CFs' activity contributed constructively to the overall broadband compound response.

The contribution of different CFs to the far-field response was assessed qualitatively by dividing the simulated AEP into two cochlear bands, spanning 3 octaves each for low and high frequencies (Figs. 5 and 6), similar to Dau, (2003). A more quantitative exploration of the model-based generation of periodic potentials (FFRs and EFRs) is presented in Fig. 8. Due to the additive nature of convolution, the AEP obtained by convolving the AN response of each individual CFs with the UR and subsequently summing them is equivalent to convolving the UR with the compound AN response, after combining the individual responses. In this analysis, the magnitude ($A_o$) and phase ($\theta$) of the spectrum at the frequency of the pure tone (300 Hz) or AM tone modulator (120 Hz) were calculated for each CF after individual convolution with the UR. The projection of each CF's response onto the total response component (at 300 or 120 Hz) was determined as $A_o \cos(\theta_{\text{total}} - \theta)$ (Fig. 8, bottom traces in gray), where $\theta_{\text{total}}$ represents the phase of the total response at the frequency of interest (Fig. 8, dark blue horizontal bands). In this detailed examination, the 2001 model showed a substantial response in both low (125–500 Hz) and high (2–10 kHz) CFs, featuring a noticeable "dip" at mid-frequencies indicative of undesired distortions stemming from the aforementioned static compressive nonlinearity. In contrast, the 2018 model displayed a broadband response, 10 dB larger at the lower frequencies and a dip at 300 Hz attributed to on-CF saturation in the IHCs. The observed decrease in magnitude at CF was not present in the EFR, as shown in Fig. 8(C) at 4 kHz. This observation is in contrast to recent modeling studies by Alamri and Jennings (2023) and Encina-Llamas *et al.* (2019), although these studies utilized SAM stimuli, unlike the RAM stimuli discussed here [see Figs. 7(A) and 7(B)]. The 2018 model further suggested that AN phase-locking to the modulation frequency is limited to higher CFs (above 3 kHz) when using RAM tones. This finding is consistent with previous modeling work indicating a predominant contribution to the EFR from ANFs tuned to frequencies above the carrier (4 kHz), often referred to as off-CF or basal ANFs (Alamri and Jennings, 2023; Encina-Llamas *et al.*, 2019). Notably, a phase shift is observed in the response of both models to the 300-Hz pure tones, resulting in a negative projection onto the total response component [Figs. 8(A) and 8(B)]. Interestingly, the

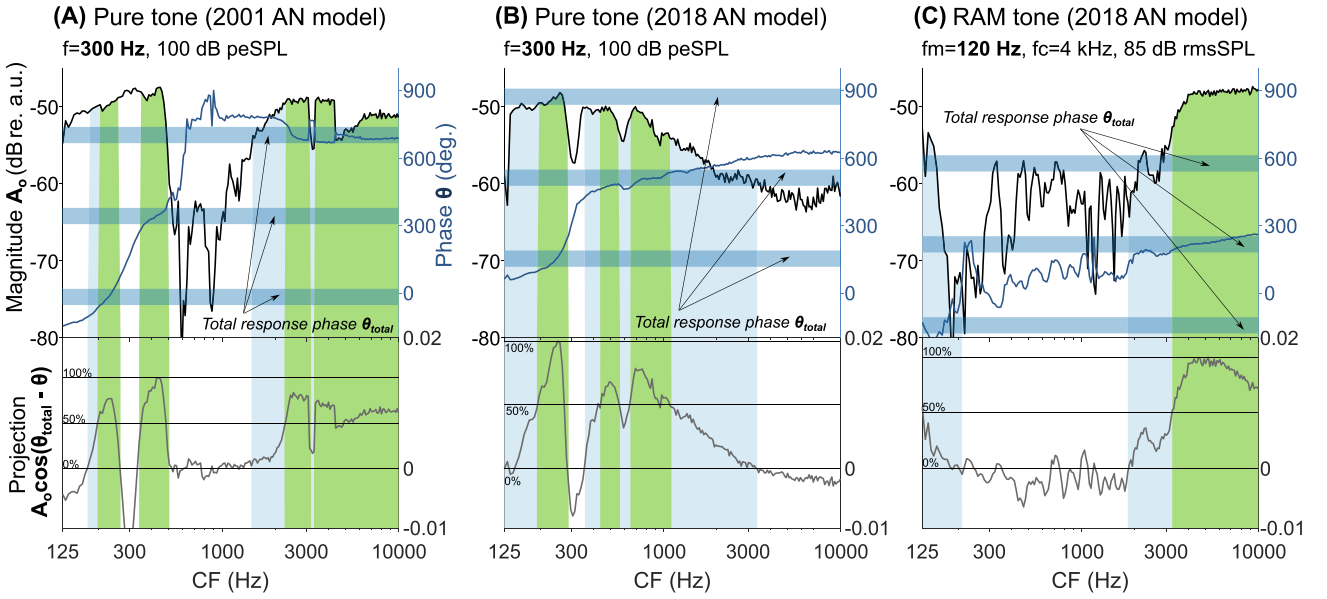


FIG. 8. (Color online) Magnitude (top left axis, black) and phase (top right axis, blue) of the AN model response of individual CFs after convolution with the UR in response to different periodic stimuli. The projections of individual CFs onto the total response component are depicted in gray below. The "total response component" refers to the amplitude and phase of the simulated AEP at the stimulation frequency (i.e., the pure tone frequency or AM tone modulation frequency). The term projection refers to the portion of the CF response that oscillates in-phase with the total response component, indicating its contribution to the overall response. Panel (A) illustrates the magnitude (top left, black), phase (top right, blue), and projection (bottom, gray) of AEPs simulated using the 2001 model at 300 Hz, corresponding to a high-level, 300-Hz tone. Panel (B) shows the corresponding simulations obtained with the 2018 model. Panel (C) displays the magnitude (top left, black), phase (top right, blue), and projections (bottom, gray) of AEPs simulated with the 2018 model categorized by CF, responding at 120 Hz to a high-level RAM tone with a carrier frequency of 4 kHz and modulation frequency of 120 Hz. The phase of the total response component is depicted with darker blue horizontal bands and is "unwrapped" to present it in multiples of 360°, eliminating any discontinuities in the plotted phase traces. CFs that contribute positively, ranging from 50 to 100% of the maximum projection, are indicated in green. In contrast, CFs with contributions between 0 and 50% of the maximum projection are shown in light blue. For visual clarity, positive contributions that were less than 10% of the maximum projection were not highlighted.

phase of the response, particularly the cochlear delay [Fig. 8(B), top blue trace, >300 Hz], significantly influences the individual CF's projection onto the total response component. With the 2001 model, the cochlear delay is challenging to discern but appears relatively short at high CFs (horizontal phase), consistent with the simulations in cats [Fig. 2(D)]. The 2018 model exhibits a larger cochlear delay, appearing less pronounced in response to the RAM tone [Fig. 8(C), top blue trace, >3 kHz] compared to the pure tone condition [Fig. 8(B), top blue trace, >300 Hz] due to the lower frequency (120 vs 300 Hz). This delay results in higher CFs (>3 kHz) having a negative projection onto the total response component, owing to a phase difference of approximately 100° with the total phase. This analysis suggests that low-frequency pure tones and high-carrier AM tones probe different sections of the cochlea (apical and basal, respectively), making them potentially valuable as complimentary diagnostic measures for IHC de-afferentation. The cochlear delay utilized in the *human* version of the AN model (Bruce *et al.*, 2018) is identical to the cochlear delay used for the *cat*. Consequently, caution is advised in interpreting projections that combine magnitude and phase. These findings underscore the importance of accurately estimating human cochlear delays, which are intricately tied to cochlear tuning characteristics, and their incorporation into computational AN models.

The present study focused on investigating responses in the "healthy" auditory system, which represents a crucial step towards better understanding the capabilities and limitations of the model in predicting AEPs in a young, normal-hearing "reference" system. In addition to studying basic stimuli, more complex stimuli, such as complex tones and speech sounds, can be employed to explore complex EFRs (Encina-Llamas *et al.*, 2019; Keshishzadeh *et al.*, 2020; Mepani *et al.*, 2021; Walton *et al.*, 2002). An intriguing application of the modeling is to systematically explore the connection between auditory pathophysiology and the generation of evoked potentials as measured in clinical settings. Previous studies have offered valuable insights in auditory pathophysiology in clinical populations using a computational AN model (Drakopoulos *et al.*, 2022; Vasilkov *et al.*, 2021; Verhulst *et al.*, 2018). Simulating the complete AEP waveform could enhance our understanding of response morphology in patients with various types of hearing loss, thereby assisting in individual diagnostics. However, accurately simulating the AN response and deriving the individual UR of an individual with hearing loss pose challenges. Since the "ground truth" of an individual's specific cochlear pathology is unknown, simulating the AN response for a listener with a hearing loss becomes inaccurate, and as a result, their individual UR cannot be obtained. Nevertheless, a general UR could be derived by averaging across many healthy participants, and this average UR could be used in

conjunction with a version of the AN model that simulates hearing loss to predict AEPs in individuals with hearing loss. Different simulated peripheral pathologies could modify the predicted potentials to varying degrees of agreement with the measured potentials.

The model framework presented in this study provides an opportunity to evaluate the accuracy of humanized AN models, which are challenging to validate through other means. Computational models of AN physiology are typically developed using animal models, where experimental data, such as single-ANF responses, are more readily available. Simulated responses are compared with these datasets, and model parameters are validated or adjusted to achieve greater physiological accuracy. However, in humans, the availability of datasets that include responses from individual ANF responses to a range of stimuli is quite limited. These datasets are mostly derived from surgical procedures, such as cochlear implant surgeries or operations on the cerebellopontine angle (Huet *et al.*, 2022). Consequently, it is difficult to verify the physiological realism of the updated parameters during the "humanization" process. This modeling framework provides an opportunity to indirectly compare results from human AN models to human measurements through non-invasive evoked potentials.

In summary, this framework may offer a valuable tool to assess early processing in humans, using both basic and more complex stimuli. It may provide a solid foundation for studying the effects of different types of hearing loss and age on AEP patterns. While not a definite standard, it is a useful tool for examining how AEPs manifest, assuming "archetypical" patterns of auditory dysfunction, wherein individual URs can be estimated and subsequently maintained as a constant for each individual. This framework also enables the exploration of related cochlear synaptopathy effects on ABR, FFR, and EFR morphology.

## V. CONCLUSION

Overall, the findings of the present study emphasize the impact of cochlear and AN processing properties on predictions of AEP and highlight the notable improvements achieved with the 2018 AN model in capturing physiological response characteristics. The model-based interpretation of the neural sources driving the generation of the FFR naturally relies on the specific characteristics of the neural processing in the AN described in the respective models. With the 2018 AN model demonstrating higher accuracy in predicting several neural features observed in physiological animal data and closer resemblance to evoked potential predictions of ABR data, it suggests a more substantial broadband contribution to the generation of high-level FFR, compared to earlier conclusions based on the previous AN modeling framework (Dau, 2003). This work may provide an interesting basis for investigating responses to more complex stimuli at both AN and retro-cochlear levels, and it offers perspectives on studying the effects of hearing loss, age, and other factors.

## ACKNOWLEDGMENTS

This work was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to declare.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Alamri, Y., and Jennings, S. G. (**2023**). "Computational modeling of the human compound action potential," J. Acoust. Soc. Am. **153**, 2376–2392.
Blackburn, C. C., and Sachs, M. B. (**1989**). "Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis," J. Neurophysiol. **62**, 1303–1329.
Bourien, J., Tang, Y., Batrel, C., Huet, A., Lenoir, M., Ladrech, S., Desmadryl, G., Nouvian, R., Puel, J.-L., and Wang, J. (**2014**). "Contribution of auditory nerve fibers to compound action potential of the auditory nerve," J. Neurophysiol. **112**, 1025–1039.
Bruce, I. C., Erfani, Y., and Zilany, M. S. A. (**2018**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites," Hear. Res. **360**, 40–54.
Bruce, I. C., Sachs, M. B., and Young, E. D. (**2003**). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," J. Acoust. Soc. Am. **113**, 369–388.
Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**, 401–417.
Carney, L. H., Li, T., and McDonough, J. M. (**2015**). "Speech coding in the brain: Representation of vowel formants by midbrain neurons tuned to sound fluctuations," eNeuro **2**, ENEURO.0004-15.2015.
Chen, J., and Jennings, S. G. (**2022**). "Temporal envelope coding of the human auditory nerve inferred from electrocochleography: Comparison with envelope following responses," J. Assoc. Res. Otolaryngol. **23**, 803–814.
Dau, T. (**2003**). "The importance of cochlear processing for the formation of auditory brainstem and frequency following responses," J. Acoust. Soc. Am. **113**, 936–950.
Dau, T., Wegner, O., Mellert, V., and Kollmeier, B. (**2000**). "Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion," J. Acoust. Soc. Am. **107**, 1530–1540.
Drakopoulos, F., Vasilkov, V., Osses Vecchi, A., Wartenberg, T., and Verhulst, S. (**2022**). "Model-based hearing-enhancement strategies for cochlear synaptopathy pathologies," Hear. Res. **424**, 108569.
Encina-Llamas, G., Dau, T., and Epp, B. (**2021**). "On the use of envelope following responses to estimate peripheral level compression in the auditory system," Sci. Rep. **11**, 6962.
Encina-Llamas, G., Harte, J. M., Dau, T., Shinn-Cunningham, B., and Epp, B. (**2019**). "Investigating the effect of cochlear synaptopathy on envelope following responses using a model of the auditory nerve," J. Assoc. Res. Otolaryngol. **20**, 363–382.
Fobel, O., and Dau, T. (**2004**). "Searching for the optimal stimulus eliciting auditory brainstem responses in humans," J. Acoust. Soc. Am. **116**, 2213–2222.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Goldberg, J. M. (**1975**). "Physiological studies of auditory nuclei of the pons," in *Auditory System. Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin), Vol. 5/2, pp. 109–144.
Goldstein, M. H., and Kiang, N. Y. S. (**1958**). "Synchrony of neural activity in electric responses evoked by transient acoustic stimuli," J. Acoust. Soc. Am. **30**, 107–114.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Harte, J., Roenne, F. M., and Dau, T. (**2013**). "Modeling human auditory evoked brainstem responses to speech syllables," J. Acoust. Soc. Am. **133**, 3429.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). "Auditory nerve model for predicting performance limits of normal and impaired listeners," Acoust. Res. Lett. Online **2**, 91–96.
Huet, A., Batrel, C., Dubernard, X., Kleiber, J. C., Desmadryl, G., Venail, F., Liberman, M. C., Nouvian, R., Puel, J.-L., and Bourien, J. (**2022**). "Peristimulus time responses predict adaptation and spontaneous firing of auditory-nerve fibers: From rodents data to humans," J. Neurosci. **42**, 2253–2267.
Ibrahim, R. A., and Bruce, I. C. (**2010**). "Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues," *The Neurophysiological Bases of Auditory Perception*, edited by E. Lopez-Poveda, A. Palmer, and R. Meddis (Springer, New York), pp. 429–438.
Jiang, Z. D., Zheng, M. S., Sun, D. K., and Liu, X. Y. (**1991**). "Brainstem auditory evoked responses from birth to adulthood: Normative data of latency and interval," Hear. Res. **54**, 67–74.
Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.
Keshishzadeh, S., Garrett, M., Vasilkov, V., and Verhulst, S. (**2020**). "The derived-band envelope following response and its sensitivity to sensorineural hearing deficits," Hear. Res. **392**, 107979.
Kiang, N. Y. (**1990**). "Curious oddments of auditory-nerve studies," Hear. Res. **49**, 1–16.
Kim, D. O., Chang, S. O., and Sirianni, J. G. (**1990**). "A population study of auditory-nerve fibers in unanesthetized decerebrate cats: Response to pure tones," J. Acoust. Soc. Am. **87**, 1648–1655.
Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Lin, T., and Guinan, J. J. (**2000**). "Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.
Lutz, B. T., Hutson, K. A., Trecca, M. C., Hamby, M., and Fitzpatrick, D. C. (**2022**). "Neural contributions to the cochlear summating potential: Spiking and dendritic components," J. Assoc. Res. Otolaryngol. **23**, 351–363.
Melcher, J. R., and Kiang, N. Y. S. (**1996**). "Generators of the brainstem auditory evoked potential in cat. III. Identified cell populations," Hear. Res. **93**, 52–71.
Mepani, A. M., Verhulst, S., Hancock, K. E., Garrett, M., Vasilkov, V., Bennett, K., de Gruttola, V., Liberman, M. C., and Maison, S. F. (**2021**). "Envelope following responses predict speech-in-noise performance in normal-hearing listeners," J. Neurophysiol. **125**, 1213–1222.
Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.
Peterson, A. J., and Heil, P. (**2018**). "A simple model of the inner-hair-cell ribbon synapse accounts for mammalian auditory-nerve-fiber spontaneous spike times," Hear. Res. **363**, 1–27.
Peterson, A. J., Irvine, D. R. F., and Heil, P. (**2014**). "A model of synaptic vesicle-pool depletion and replenishment can account for the interspike interval distributions and nonrenewal properties of spontaneous spike trains of auditory-nerve fibers," J. Neurosci. **34**, 15097–15109.
Picton, T. W., Stapells, D. R., and Campbell, K. B. (**1981**). "Auditory evoked potentials from the human cochlea and brainstem," J. Otolaryngol. Suppl. **9**, 1–41.
Prosser, S., and Arslan, E. (**1987**). "Prediction of auditory brainstem wave V latency as a diagnostic tool of sensorineural hearing loss," Int. J. Audiol. **26**, 179–187.
Rønne, F. M., Dau, T., Harte, J., and Elberling, C. (**2012**). "Modeling auditory evoked brainstem responses to transient stimuli," J. Acoust. Soc. Am. **131**, 3903–3913.
Scherg, M., and Von Cramon, D. (**1985**). "A new interpretation of the generators of BAEP waves I–V: Results of a spatio-temporal dipole model," Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sect. **62**, 290–299.
Serpanos, Y. C., O'Mulley, H., and Gravel, J. S. (**1997**). "The relationship between loudness intensity functions and the click-ABR wave V latency," Ear Hear. **18**, 409–419.
Shaheen, L. A., Valero, M. D., and Liberman, M. C. (**2015**). "Towards a diagnosis of cochlear neuropathy with envelope following responses," J. Assoc. Res. Otolaryngol. **16**, 727–745.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2001**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Spoendlin, H., and Schrott, A. (**1989**). "Analysis of the human auditory nerve," Hear. Res. **43**, 25–38.
Strelcyk, O., Christoforidis, D., and Dau, T. (**2009**). "Relation between derived-band auditory brainstem response latencies and behavioral frequency selectivity," J. Acoust. Soc. Am. **126**, 1878–1888.
Vasilkov, V., Garrett, M., Mauermann, M., and Verhulst, S. (**2021**). "Enhancing the sensitivity of the envelope-following response for cochlear synaptopathy screening in humans: The role of stimulus envelope," Hear. Res. **400**, 108132.
Verhulst, S., Altoè, A., and Vasilkov, V. (**2018**). "Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss," Hear. Res. **360**, 55–75.
Verhulst, S., Bharadwaj, H. M., Mehraei, G., Shera, C. A., and Shinn-Cunningham, B. G. (**2015**). "Functional modeling of the human auditory brainstem response to broadband stimulation," J. Acoust. Soc. Am. **138**, 1637–1659.
Walton, J. P., Simon, H., and Frisina, R. D. (**2002**). "Age-related alterations in the neural coding of envelope periodicities," J. Neurophysiol. **88**, 565–578.
Wegner, O., and Dau, T. (**2002**). "Frequency specificity of chirp-evoked auditory brainstem responses," J. Acoust. Soc. Am. **111**, 1318–1329.
Yin, T. C., and Chan, J. C. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.
Zhang, X., and Carney, L. H. (**2005**). "Analysis of models for the synapse between the inner hair cell and the auditory nerve," J. Acoust. Soc. Am. **118**, 1540–1553.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.
Zilany, M. S. A., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. **120**, 1446–1466.
Zilany, M. S. A., and Bruce, I. C. (**2007**). "Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats," J. Acoust. Soc. Am. **122**, 402–417.
Zilany, M. S. A., Bruce, I. C., and Carney, L. H. (**2014**). "Updated parameters and expanded simulation options for a model of the auditory periphery," J. Acoust. Soc. Am. **135**, 283–286.
Zilany, M. S. A., Bruce, I. C., Nelson, P. C., and Carney, L. H. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," J. Acoust. Soc. Am. **126**, 2390–2412.