



OPEN ACCESS

EDITED BY

Alessia Paglialonga,
Institute of Electronics, Information Engineering
and Telecommunications (IEIT), Italy

REVIEWED BY

Tobias Weissgerber,
University Hospital Frankfurt, Germany
Etienne Gaudrain,
INSERM U1028 Centre de Recherche en
Neurosciences de Lyon, France

*CORRESPONDENCE

Waldo Nogueira
✉ NogueiraVazquez.Waldo@mh-hannover.de

RECEIVED 02 May 2022

ACCEPTED 15 February 2023

PUBLISHED 09 March 2023

CITATION

Alvarez F, Kipping D and Nogueira W (2023) A
computational model to simulate spectral
modulation and speech perception
experiments of cochlear implant users.
Front. Neuroinform. 17:934472.
doi: 10.3389/fninf.2023.934472

COPYRIGHT

© 2023 Alvarez, Kipping and Nogueira. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

A computational model to simulate spectral modulation and speech perception experiments of cochlear implant users

Franklin Alvarez^{1,2}, Daniel Kipping^{1,2} and Waldo Nogueira^{1,2*}

¹Medizinische Hochschule Hannover, Hannover, Germany, ²Cluster of Excellence "Hearing4All", Hannover, Germany

Speech understanding in cochlear implant (CI) users presents large intersubject variability that may be related to different aspects of the peripheral auditory system, such as the electrode–nerve interface and neural health conditions. This variability makes it more challenging to proof differences in performance between different CI sound coding strategies in regular clinical studies, nevertheless, computational models can be helpful to assess the speech performance of CI users in an environment where all these physiological aspects can be controlled. In this study, differences in performance between three variants of the HiRes Fidelity 120 (F120) sound coding strategy are studied with a computational model. The computational model consists of (i) a processing stage with the sound coding strategy, (ii) a three-dimensional electrode-nerve interface that accounts for auditory nerve fiber (ANF) degeneration, (iii) a population of phenomenological ANF models, and (iv) a feature extractor algorithm to obtain the internal representation (IR) of the neural activity. As the back-end, the simulation framework for auditory discrimination experiments (FADE) was chosen. Two experiments relevant to speech understanding were performed: one related to spectral modulation threshold (SMT), and the other one related to speech reception threshold (SRT). These experiments included three different neural health conditions (healthy ANFs, and moderate and severe ANF degeneration). The F120 was configured to use sequential stimulation (F120-S), and simultaneous stimulation with two (F120-P) and three (F120-T) simultaneously active channels. Simultaneous stimulation causes electric interaction that smears the spectrotemporal information transmitted to the ANFs, and it has been hypothesized to lead to even worse information transmission in poor neural health conditions. In general, worse neural health conditions led to worse predicted performance; nevertheless, the detriment was small compared to clinical data. Results in SRT experiments indicated that performance with simultaneous stimulation, especially F120-T, were more affected by neural degeneration than with sequential stimulation. Results in SMT experiments showed no significant difference in performance. Although the proposed model in its current state is able to perform SMT and SRT experiments, it is not reliable to predict real CI users' performance yet. Nevertheless, improvements related to the ANF model, feature extraction, and predictor algorithm are discussed.

KEYWORDS

computational model, cochlear implant, neural health, sound coding strategies, speech-in-noise recognition, spectral modulation detection, speech understanding prediction

1. Introduction

People diagnosed with severe or profound sensorineural hearing loss that keep some healthy auditory nerve fibers (ANFs) are good candidates to receive a cochlear implant (CI) and recover to some extent their sense of hearing. A CI consists of an electrode array implanted in the cochlea, and a wearable sound processor usually located behind the ear. The sound processor is responsible for converting acoustic signals into electric stimulation patterns that are delivered to the ANFs *via* the intracochlear electrodes (Wouters et al., 2015). In many auditory tasks, there is a big gap in performance between normal hearing (NH) and CI listeners (Nelson et al., 2003; Nelson and Jin, 2004). Electric stimulation has its limitations to convey the necessary information for the proper coding of sounds in the auditory system (Moore, 2003). To reduce this gap, researchers are dedicated to find better CI sound coding strategies (Nogueira et al., 2005, 2009; Landsberger and Srinivasan, 2009; Dillon et al., 2016; Langner et al., 2020a; Gajeci and Nogueira, 2021), but the evaluation of the potential benefits of new ideas usually requires extensive testing procedures with implanted volunteers. In addition, there is high variability in the performance among CI users (Moberly et al., 2016), which makes it more difficult to generalize from the results.

CI sound coding strategies using current steering aim at providing an increased number of stimulation places in the implanted cochlea (Landsberger and Srinivasan, 2009; Nogueira et al., 2009). The general idea is to create virtual channels by “steering” the electrical field between two adjacent electrodes, balancing their output current at different ratios. The commercial sound coding strategy HiRes with Fidelity120 (F120), from Advanced Bionics, offers up to 120 virtual channels using 16 electrodes because every electrode pair is able to steer the electrical field to eight different locations. Furthermore, power savings can be achieved by stimulating various virtual channels simultaneously. Simultaneous stimulation allows to increase the pulse duration and consequently decrease the maximum current needed (Langner et al., 2017). The drawback is that simultaneous stimulation produces electric interaction that causes spectral smearing across channels, which also causes temporal smearing since temporal modulations may be reduced (Nogueira et al., 2021). The balance between power savings and CI users’ performance was investigated by Langner et al. (2017) using three variations of the F120. Sequential stimulation (F120-S), where one virtual channel was active at a time, was compared to paired (F120-P) and triplet (F120-T) stimulation, where two and three virtual channels were active at the same time, respectively. They found out that the channel interaction that occurs with the simultaneous stimulation in F120-P and F120-T has a negative impact on performance, with F120-T obtaining the worst score. Nevertheless, high inter-subject variability was found in speech intelligibility and spectral modulation detection threshold.

It has been shown that peripheral aspects such as neural health condition (Nadol, 1997), insertion depth and position of the electrode array (Dorman et al., 1997), along with more central aspects such as neural plasticity (Han et al., 2019) may account for an important part of the inter-subject variability observed in CI users. However, it is not possible to estimate the degree of

ANF degeneration without invasive methods unless the individual is already implanted with a CI (Prado-Gutierrez et al., 2006; Ramekers et al., 2014; Imsiecke et al., 2021; Langner et al., 2021). Langner et al. (2021) investigated the hypothesis that individuals with good neural health and electrode positioning will show a lower difference in performance when using simultaneous stimulation strategies (F120-P and F120-T) compared to sequential stimulation (F120-S). Healthy conditions lead to lower focused thresholds and less channel interaction between virtual channels; therefore, healthy neural conditions could lead to less detriment in performance when comparing sequential and simultaneous stimulation. The performance was evaluated using the Hochmair–Schulz–Moser (HSM) sentence test (Hochmair-Desoyer et al., 1997) at a signal-to-noise ratio (SNR) where participants roughly understood 50% of the words with F120-S. The results showed no correlation between any measure intended to estimate the neural health and difference in performance, arguably, because of the small number of individuals measured. On the contrary, computational models that simulate the electrode–nerve interface in CI users can assess the relation between neural health and performance. These computational models can isolate the parameter of study to remove the inter-subject variability, i.e., nerve count, nerve degeneration, electrode position, or insertion depth.

Computational models of the electrode–nerve interface for CIs have been proposed at different levels of complexity. Fredelake and Hohmann (2012) presented a one-dimensional interface model where the ANFs are equally distributed along a cochlear axis with the electrode array positioned in the center of this ANF population with equidistant electrodes. The spatial spread of stimulation was calculated depending on the distance between electrodes and ANFs with an exponential decay function. Neural health conditions with this model were assessed by changing the ANF density while maintaining the total neural activity constant. Lower ANF density requires higher current levels; therefore, the excitation from a single electrode reaches further ANFs in the cochlear axis causing channel interaction and spectral smearing. However, this electrode–nerve interface is very limited when representing physical aspects that occur in real implantation. From clinical imaging data, a patient-specific three-dimensional model of the implanted cochlea can be constructed (Rattay et al., 2001; Stadler and Leijon, 2009; Kalkman et al., 2014; Malherbe et al., 2016; Nogueira et al., 2016; Heshmat et al., 2020, 2021; Croner et al., 2022). These models fit a population of ANFs (type 1 spiral ganglion neurons) that extend from the organ of Corti to the central axons. Also, it is possible to control the positioning of the electrode array inside the scala tympani. The voltage spread produced by the electric stimulation from the electrode array can be calculated using a homogeneous model of the extracellular medium (Rattay et al., 2001; Litvak et al., 2007a; Nogueira et al., 2016), or using a finite element method (FEM) to account for the different electrical properties of all structures between the stimulating electrode and the ANF population (Nogueira et al., 2016). Such an electrode–nerve interface can be coupled with an ANF model capable of simulating action potentials (also called spikes) from the electrical stimulation (Ashida and Nogueira, 2018).

Regarding the ANF models, there are two different approaches. The “physiological” approach aims to simulate processes on

a microscopic level. An example is the multi-compartment Hodgkin–Huxley model (Rattay, 1999; Rattay et al., 2001; Smit et al., 2008) that offers a very precise electrical behavior of the ANF segments when transmitting the action potentials throughout the peripheral axon, the soma, and the central axon. The drawback of this approach is the high demand for computational resources. The “phenomenological” approach tries to reproduce the effective outcome without detailed simulation of the involved intermediate processes. The spike generation algorithm does not consider how the spike travels through the nervous system, hence there is no geometric information involved. Some models depend completely on a probabilistic function (Bruce et al., 1999), while others are based on a leaky integrate-and-fire electrical circuit where the membrane voltage is calculated for every time step, and when it reaches a threshold, the ANF produces an action potential (Hamacher, 2004; Joshi et al., 2017). The membrane voltage depends on many other parameters like feedback currents, refractory periods, and membrane noise to introduce stochasticity. These parameters can be adjusted to fit data measured in humans or animal models to account for physiological aspects. The output of an ANF model is the “spike train”, which consists of a binary array indicating the time frames where an action potential (spike) is produced. With a population of ANFs, it is possible to integrate the spike trains, in time and cochlear place, to obtain features that are representations of sound at higher levels in the auditory system. Integration allows to reduce the amount of data while preserving the information that reaches, for example, a speech recognition algorithm (Fredelake and Hohmann, 2012; Jürgens et al., 2018).

The simulation framework for auditory discrimination experiments (FADE) (Schädler et al., 2016) is a computational tool capable of performing speech recognition tasks and psychoacoustic experiments simulating human performance. Originally, FADE was used to simulate the performance of NH and hearing aided people (Kollmeier et al., 2016; Schädler et al., 2018). Then, a CI sound coding strategy and a CI auditory model were incorporated into FADE to perform simulations of speech reception thresholds (SRTs) using data from different CI users (Jürgens et al., 2018). The SRT is defined as the signal-to-noise ratio (SNR) where 50% of the words in a sentence are correctly identified (Wagener et al., 1999) and it is a direct indicator of the CI user performance in speech understanding. However, Jürgens et al. (2018) used the same peripheral auditory model as Fredelake and Hohmann (2012), which is a simplified one-dimensional representation of the electrode–nerve interface. The incorporation of a more complex peripheral auditory model with a three-dimensional representation of the electrode–nerve interface should turn FADE into a powerful framework to assess studies related to neural health conditions in CI users. It can also be useful to assess the benefits of novel sound coding strategies. Objective instrumental measures commonly used for this purpose rely on vocoders to simulate the degraded sound delivered by the CI (Chen and Loizou, 2011; Santos et al., 2013; El Boghdady et al., 2016), not accounting for physiological aspects of the implantation.

Performance with a CI may be also predicted using simpler behavioral measurements than the SRT. The spectral modulation threshold (SMT) is defined as the ripple depth in dBs at which 79.4% of spectral rippled noise is differentiated from flat noise. SMT

has been used alongside speech recognition experiments because it is a good indicator of how well the spectral cues in speech signals were perceived (Litvak et al., 2007b; Langner et al., 2017). It was used by Langner et al. (2017) as an indicator of how these spectral cues are affected by the channel interaction occurring in simultaneous stimulation (F120-P and F120-T), compared to sequential stimulation (F120-S). Their results showed similar performance between F120-S and F120-P but a clear lowering of performance with F120-T.

In this study, a computational model that simulates the performance of real CI users in SRT and SMT experiments is presented. The goal is to show the effects of parallel stimulation and neural degeneration in CI outcome performance. This model was tested with the three sound coding strategies used by Langner et al. (2021) (F120-S, F120-P and F120-T), and the hypothesis that channel interaction affects individuals with poorer neural health conditions to a larger extent is assessed. In the next section, the different parts composing this computational model and how the SRT and SMT experiments were implemented with FADE are described. A further section presents the results obtained, and the last section contains the discussion and conclusions of this study.

2. Materials and methods

2.1. The computational model

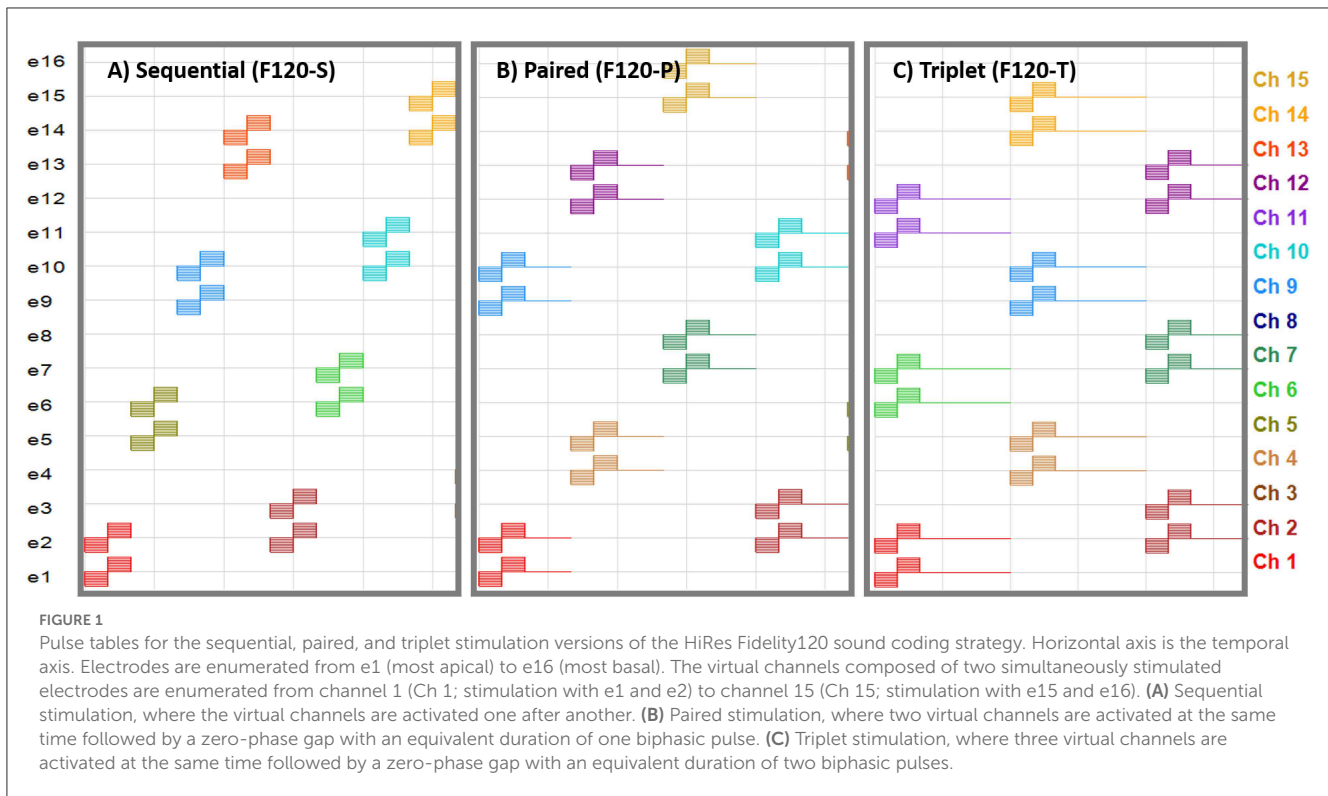
The proposed computational model consists of (i) a “front-end” containing the CI sound coding strategy, the peripheral auditory model with a three-dimensional electrode–nerve interface, and a feature extraction algorithm; (ii) a “back-end” with a hidden Markov model (HMM) already incorporated in the framework FADE.

2.1.1. Front-end

2.1.1.1. Cochlear implant sound coding strategy

The software BEPS+ from Advanced Bionics was used to create the pulse tables for the F120-S, F120-P, and F120-T sound coding strategies. The pulse tables are defined as the sequence of electrical pulses to create one cycle of stimulation. Figure 1 shows partial pulse tables corresponding to these strategies. A pulse consists of a cathodic-leading biphasic pulse. The electrodes are enumerated from the most apical to the most basal and each virtual channel composed of two simultaneously stimulated electrodes are depicted with a color code. The pulse phase duration was set to 18 μ s and the pulse rate across virtual channels was kept constant at 1,852 pps by adding a gap between subsequent pairs, or triplets, of stimulating virtual channels for F120-P and F120-T, respectively.

The HiRes implantable cochlear stimulator (ICS) from Advanced Bionics was used to transform audio signals into electrograms. The audio signal was calibrated to -49 dB full scale [dB_{F5}], corresponding with an audio signal at 65 dB sound pressure level [dB_{SPL}] captured by the microphone of the CI device. At this value, the signal level was close to the knee point of the adaptive gain control of the CI sound processor.



Biphasic pulses of the electrodiagrams were resampled to 1 MHz to guarantee equal anodic and cathodic phases. This sample rate was also needed in the implementation of the peripheral auditory model.

2.1.1.2. Peripheral auditory model

Figure 2 shows the composition of the proposed peripheral auditory model. The electrodiagrams obtained from the sound coding strategy were transformed to obtain the voltage spread based on a three-dimensional electrode–nerve interface model embedded in a homogeneous medium. The amount of stimulation at every ANF was obtained from this voltage spread and the times when action potentials are elicited in every ANF (spike trains) were simulated with an active nerve fiber model. The spike activity is defined as the collection of spike trains produced in an ANF population.

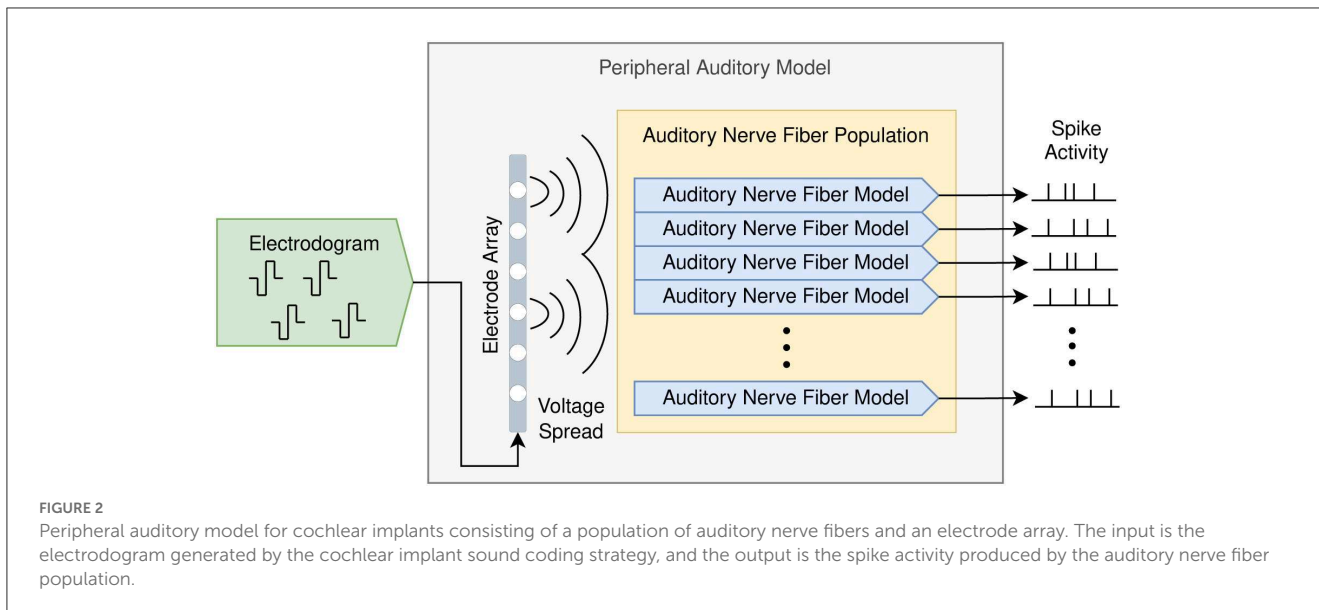
The electrode–nerve interface used in the proposed model was based on the cochlea model presented in Nogueira et al. (2016). Cochlear geometry, electrode location, and position of the ANF population were taken from a generic version of their model. The number of ANFs was increased from 7,000 to 9,001 and distributed along 900° of insertion angle from base to apex (two turns and a half) with a separation of 0.1°. The ANFs were indexed in order from the base of the cochlea (high frequencies) to the apex of the cochlea (low frequencies). Another adjustment was done to the electrode array. Nogueira et al. (2016) modeled an electrode array of 22 electrodes; therefore, the electrodes 21, 19, 17, 15, 13, and 11 were removed to obtain the 16 electrodes present in advanced bionics CIs. The resulting electrode–nerve interface is shown in Figure 3.

The morphology of the ANFs was modeled after the myelinated fibers presented in Ashida and Nogueira (2018), which is a simplified representation consisting of segments with a constant internodal length (L_i) equal to 200 μm that extends from the location of the peripheral terminal toward the cochlear nerve. In this morphological model, there is no differentiation between the peripheral axon, central axon, or the soma. As in Ashida and Nogueira (2018), the electric stimulation in the myelinated model was calculated at the nodes that join together two adjacent segments. The voltage produced by the stimulation current I_n coming out of the electrode n was calculated for every node a of every fiber f as shown in Equation (1).

$$U_{nfa} = \frac{\rho_{\text{ext}} I_n}{4\pi d_{nfa}} \tag{1}$$

The extracellular resistivity of the homogeneous medium (ρ_{ext}) was set to 3.0 Ωm as in Ashida and Nogueira (2018). The variable d_{nfa} is the distance between the electrode n and the node a . This approach results in a voltage spread inversely proportional to the distance d_{nfa} (Litvak et al., 2007a; Nogueira et al., 2016).

The activation function has been proposed by Rattay (1999) to approximate the amount of functional electrical stimulation over an ANF. In this model, it was calculated as shown in Equation (2). The activation function in a node a depends on its external voltage (U_{nfa}), and the external voltage on its adjacent nodes ($a-1$ toward the periphery and $a+1$ toward the central neural system). The axon internal resistance (R_i) was obtained as “ $R_i = 4L_i r / \pi D^2$ ”. The axon diameter (D) was set to 2.0 μm , and the axial resistivity (r) to 1.0 Ωm , as mentioned by Ashida and Nogueira (2018). Notice that the activation function in this study has units in Amperes [A]



instead of volts per second [V/s] as originally defined by Rattay (1999). This is because the membrane capacitance is not included in Equation (2); however, this membrane capacitance is taken into account in a further stage of the model.

$$A_{nfa} = \frac{U_{nfa-1} - U_{nfa}}{R_i} + \frac{U_{nfa+1} - U_{nfa}}{R_i}. \quad (2)$$

To simulate the spikes generated by each ANF, the neuron model of Joshi et al. (2017) was implemented. This is a “phenomenological” model that represents the peripheral and central axons as two independent adaptive integrate-and-fire circuits that are coupled together by a logical “OR” gate (see Figure 1 in Joshi et al., 2017). This phenomenological model does not convey any geometric information such as the distance between the stimulating electrode and the ANF. Therefore, the induced current I (called stimulation current in Joshi et al., 2017) was adjusted according to the activation function obtained from the electrode–nerve interface model as shown in Equation (3).

$$I = M_C \sum_{n=1}^N A_{nfa_{max}}. \quad (3)$$

Notice that the activation function in Equation (2) has a value for every node in an ANF. To simplify the implementation, only the node (a_{max}) with the maximum absolute value of the activation function was taken into account to compute the induced current I . This is based on the fact that this is the node with the highest probability to produce a spike. In addition, a modeling factor (M_C) that allowed to calibrate the peripheral auditory model was added. It was adjusted to reproduce approximately the same spike count reported by Joshi et al. (2017) given different stimulation current levels.

The model of Joshi et al. (2017) assumes that the peripheral and central circuits share the same induced current (I), but they respond differently to the positive (anodic; I^+) and negative (cathodic;

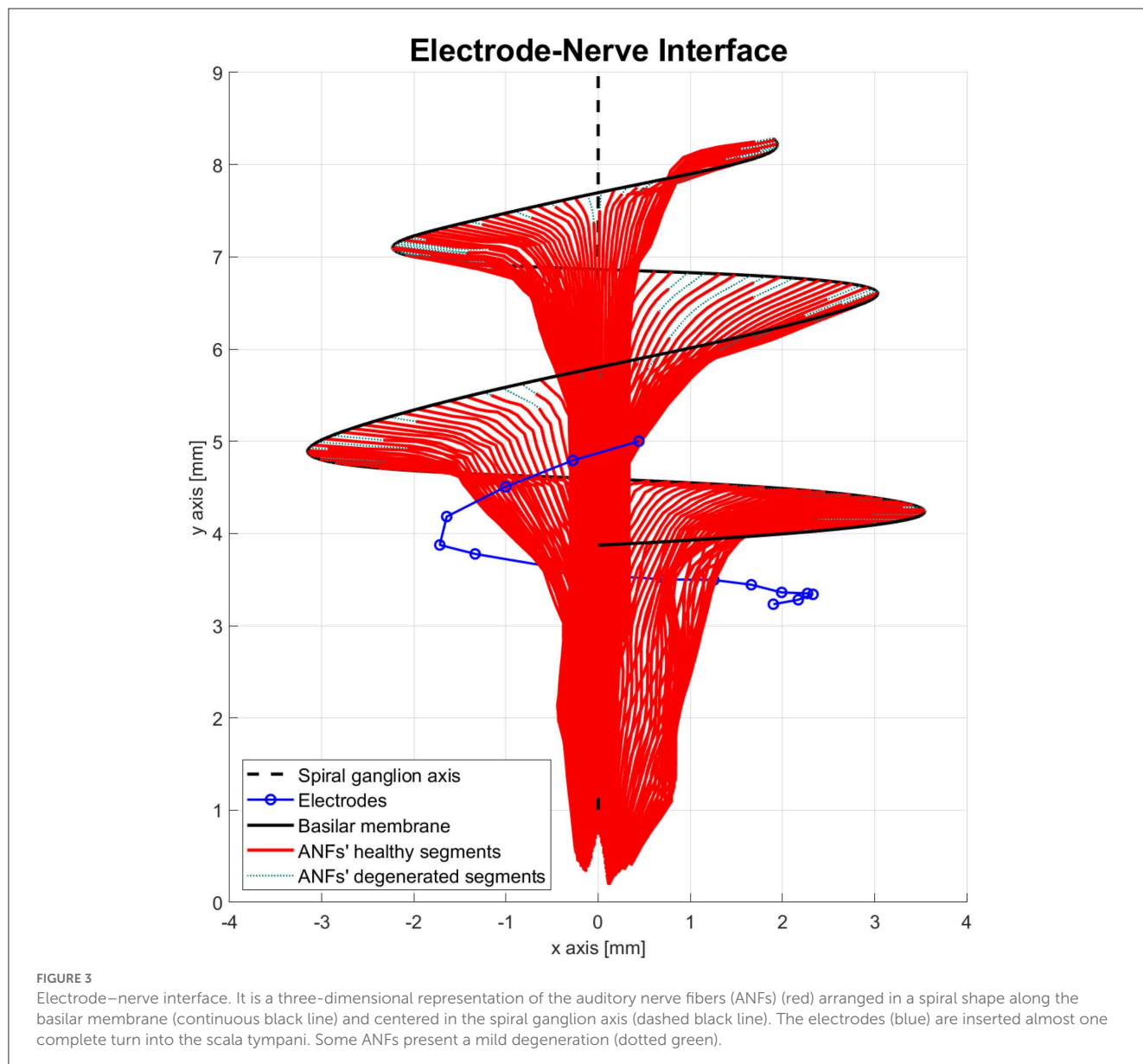
I^-) phases of the biphasic pulses. Therefore, in this study, the peripheral axon circuit is referred to as cathodic-excitatory while the central axon circuit as anodic-excitatory. The circuit specific induced current (I_{stim}) was obtained with Equation (4), where the inhibitory compression (β) was set to 0.75.

$$I_{stim} = \begin{cases} -(I^- + \beta I^+) & \text{Cathodic-excitatory circuit.} \\ I^+ + \beta I^- & \text{Anodic-excitatory circuit.} \end{cases} \quad (4)$$

The membrane voltage (V) for both circuits is calculated with Equation (5), where the membrane capacitance (C) takes different values for the cathodic-excitatory (856.96 nF) and the anodic-excitatory (1772.4 nF) circuit, $h(V)$, is a passive filter dependent on membrane voltage, I_{sub} and I_{supra} are internal subthreshold and suprathreshold adaptation currents, and I_{noise} is a noise current source with a Gaussian spectral shape that introduces stochastic behavior into the spike trains. The passive filtering, the evolution of the adaption currents and the noise have their own function and can be found in the publication of Joshi et al. (2017). Whenever the membrane voltage of the cathodic-excitatory or the anodic-excitatory circuit reached a threshold, a spike was generated and the ANF entered in an absolute refractory period (ARP) of 500 μ s. During the ARP, neither the cathodic- nor anodic-excitatory circuit could produce a spike.

$$C \frac{dV}{dt} = h(V) - I_{sub} - I_{supra} + I_{noise} + I_{stim}. \quad (5)$$

Another important feature of the proposed peripheral auditory model is the representation of different neural health conditions. A degeneration index (α_f) was assigned to every ANF, which was a natural number from 0 to 20, indicating how many segments were removed from its modeled morphology. The segments were always removed from the most peripheral part resembling the dendritic degeneration that occurs when the inner hair cells in the basilar membrane are damaged (Spendlin and Schrott, 1988;



Nadol, 1997). The nerve degeneration was limited to 20 segments because, beyond this point, the amount of stimulation current required to elicit a spike was excessive compared to real CI users. Figure 4 shows how nerve degeneration was implemented in the proposed electrode–nerve interface.

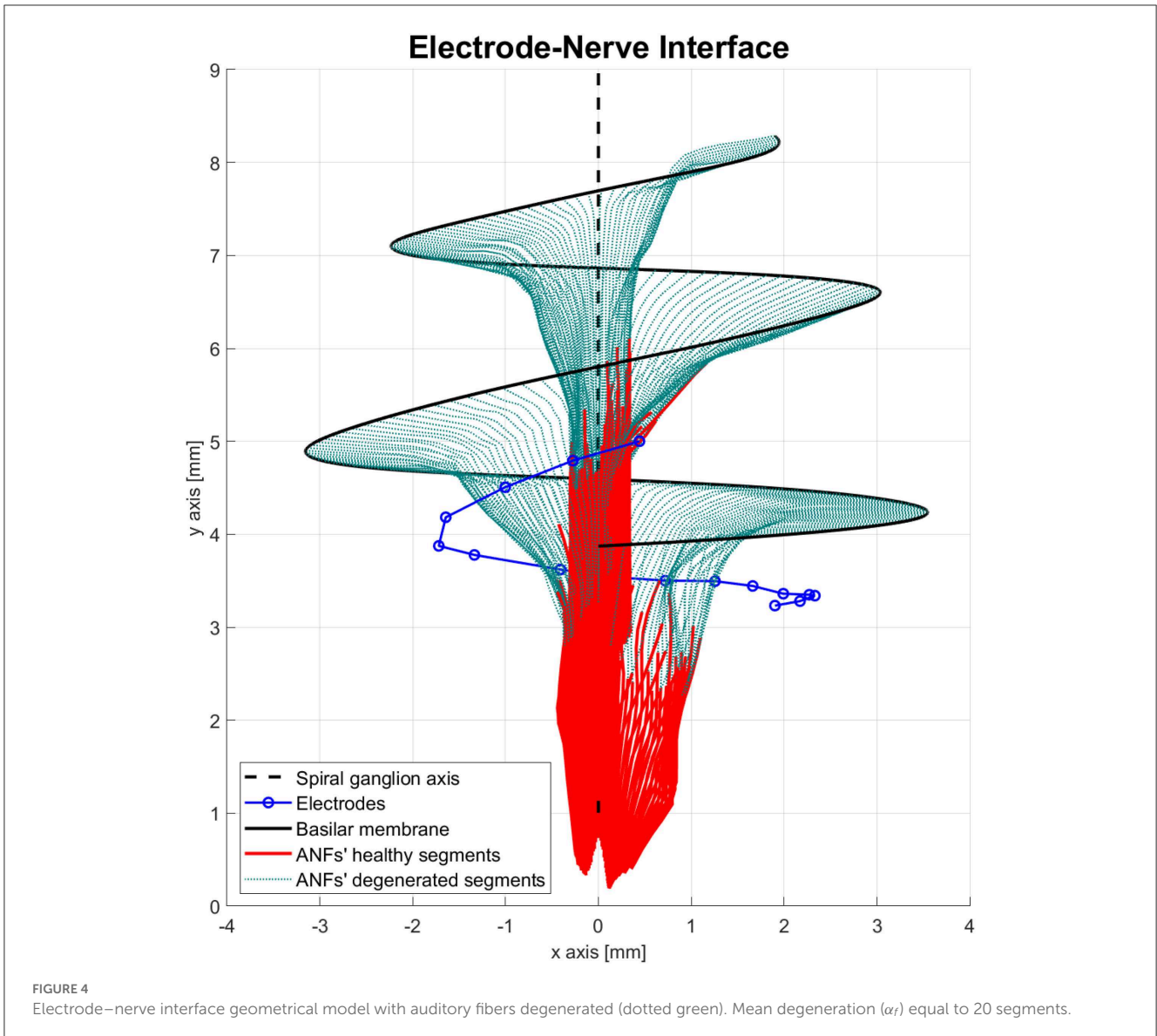
It is worth mentioning that removing segments may result in situations where the degenerated part surpasses the physical location of the soma, which in real spiral ganglion neurons would be somewhere between the seventh and the twelfth segment. Nevertheless, degeneration of the peripheral axon could also be modeled as the loss of myelin sheets, or by reducing its diameter (Heshmat et al., 2020, 2021; Croner et al., 2022). The effects of this type of degeneration would be that nodes in the central axon will be the ones that produce a spike. Therefore, removing segments in our model was used to investigate excitation at most central locations, rather than to represent the real physical degeneration.

Because it is unknown how the current flows in the most peripheral nodes after degeneration, it was decided to discard them from the activation function calculation. In this regard, Rattay (1999) proposed to remove the first term in Equation (2); nevertheless, in degenerated ANFs, following this proposal could result in a rise of the activation function despite the worst neural health condition and this effect was undesired in our model.

2.1.1.3. Internal representation as features

The feature extraction algorithm was based on the internal representation (IR) presented by Fredelake and Hohmann (2012), which accounts for more central processes in the auditory pathway. The IR consists of a spatial and a temporal integration of the spike activity produced by the ANF population. The first step was to downsample the spike activity to a sample rate of 10 kHz.

To perform the spatial integration, the ANFs were grouped resembling the auditory filters described in Moore (2003). It is



mentioned that a maximum of 39 independent auditory filters can be formed at the same time to code sound in NH people, but it is also mentioned that the effective number of channels for CI users could be reduced depending on the number of electrodes implanted. Therefore, in the computational model, the number of auditory filters used was limited between 16 (number of electrodes) and 39. Lengths of the auditory filters ranged between 1.1 and 2.6 mm in the 42 mm basilar membrane of the modeled cochlea. To obtain the number and distribution of these auditory filters along the basilar membrane, adjacent ANFs were grouped by their most likely stimulating electrode (highest absolute value of the activation function) to form auditory filters. If an auditory filter was below the minimum size (1.1 mm), its fibers were merged with the adjacent auditory filter toward the apex. In case an auditory filter size exceeded the maximum value (2.6 mm), its most basal ANFs were used to form a new auditory filter of maximum size while the remaining ANFs were used to form a different auditory filter. Once the auditory filters were formed, the spike trains of their

corresponding ANFs were added together to obtain the spike group activity (S_g), where g was the auditory filter index.

The next step was to integrate this spike group activity across time. For each group, the signal was low pass filtered as shown in Equation (6), where $F_g(k)$ is the filtered spike group activity, k is the time frame index, f_s is the sample frequency equal to 10.0 kHz, τ_{LP} is the time constant of the filter set to 1 ms, and the operator “*” denotes a convolution.

$$F_g(k) = S_g(k) * \exp\left(-\left(\frac{k}{\sqrt{2}f_s\tau_{LP}}\right)^2\right). \quad (6)$$

From this point, a forward masking effect is implemented. A masker signal Z_g was derived from the filtered spike group activity using a recursive low pass filter (see Section 2.3 in Fredelake and Hohmann, 2012). This masker signal increases exponentially with onsets in the spike group activity and decreases exponentially with

the offsets. The IR is finally the maximum value between the masker signal and the filtered spike group activity. The whole forward masking effect is taken from [Fredelake and Hohmann \(2012\)](#) and it is not detailed in this study. A visual representation is found in Figure 4 of [Fredelake and Hohmann \(2012\)](#).

To meet the requirements of the back-end, the IR was further downsampled to a sample rate of 100 Hz using a moving average low pass filter to diminish the effects of aliasing.

2.1.2. Back-end

The computational model used the framework FADE as the back-end. As a predictor algorithm, FADE uses an HMM that represents the target stimulus with an eight-state Markov chain and a one-component Gaussian mixture model (GMM) to learn, and subsequently predict, the features (each auditory filter in the IR) ([Schädler et al., 2016](#)). FADE counts with different “ready-to-use” experiment templates, processing algorithms, and feature extraction algorithms that are intended to predict NH and hearing aid performances ([Kollmeier et al., 2016](#); [Schädler et al., 2016, 2018](#)). Hence, two new experiment templates (described in detail in Section 2.4) were developed to perform SRT and SMT experiments. The tasks handled by these new templates are as follows:

1. The generation of the stimulus audio files composing the training and testing corpus.
2. The generation of the electrodiagrams from these audio files using the respective CI sound coding strategy as the processing algorithm.
3. The generation of the stimuli' IRs using the proposed peripheral auditory model as the feature extraction algorithm.
4. The training of the HMM with the IRs obtained from the training corpus.
5. The predictions over the IRs of the testing corpus with the trained HMM.
6. The evaluation of the performance of the HMM.

In the evaluation stage, FADE generated a file with the score obtained at different training conditions (dB_{SNR} in SRT experiments and $\text{dB}_{\text{contrast}}$ in SMT experiments). Scores were represented as data points in a scatter plot and a non-linear regression to a psychometric function was performed. This psychometric function is defined in Equation (7), where p_{chance} is the lower horizontal asymptote of the function representing the probability of getting a correct answer with random predictions, p_{max} is the upper asymptote of the function representing the predicted performance in ideal conditions, p_{range} is the difference between the upper and lower asymptotes, s is the slope, or growth rate, at the inflection point of the psychometric function, and x_0 is the offset of the inflection point in the x-axis (dB_{SNR} or $\text{dB}_{\text{contrast}}$). The regression was performed with the MATLAB function “*fitnlm*”. The coefficient of determination R^2 was obtained in every experiment, which is a reference of how well the scattered data was represented by the regressed psychometric function.

$$Ps(x) = p_{\text{chance}} + \frac{p_{\text{range}}}{1 + e^{-s(x-x_0)}}. \quad (7)$$

2.2. Fitting and calibration

The fitting procedure in CI users consists of the adjustment of the stimulation levels of each electrode in the array, or virtual channels in the case of current steering strategies such as F120. Each electrode, or virtual channel, stimulates at levels between threshold (T) and most comfortable loudness (MCL) that are unique for every CI user. By default, the advanced bionics device sets T to 10% of the MCL level resulting in a 20 dB dynamic range. Stimulation levels with the advanced bionics device are given in clinical units (CU), which are integer values from 1 to 471. The equivalent output current (I_n) [μA] was obtained with Equation (8), where T_p is the pulse duration in μs (18 μs in this study), I_{max} is the maximum output current equal to 2,040 μA , and T_{max} is the maximum pulse duration equal to 229 μs ([Advanced Bionics, 2020](#)).

$$I_n = \frac{CU}{6000} \frac{I_{\text{max}} T_{\text{max}}}{T_p}. \quad (8)$$

The process of fitting requires a closed feedback loop, where the CI user indicates the loudness perceived to an audiologist. This loop is virtually closed in the computational model by measuring the spike activity produced by electric pulse trains at different levels of stimulation (from 1 to 471 CU in steps of 30 CU) based on the assumption that the loudness perception is closely related to the neural activity produced by the ANFs ([McKay and McDermott, 1998](#); [McKay et al., 2001](#)).

The fitting stimulus used was a cathodic-leading biphasic pulse train with a pulse duration of 18 μs and a periodicity of 540 μs (resulting approximately in 1,852 pps) was consistent with the experimental parameters. The pulse train had a duration of 200 ms with 10 ms of leading and preceding silence. Because the periodicity of this fitting stimulus is just above the ARP, it was expected that ANFs close to the electrodes “fired” with every biphasic pulse of the pulse train.

For each virtual channel, a group of 858 ANFs with the highest absolute activation function was selected. This number corresponds to the number of fibers found in approximately 4 mm section of the modeled basilar membrane, although the selected ANFs were not constrained to be adjacent to each other. The MCL was defined as the CU value where each biphasic pulse elicited a spike in the selected ANF group with a probability of 75%. A similar assumption was used by [Kalkman et al. \(2014\)](#). T level was set to the 10% of the MCL level.

The calibration of the peripheral auditory model refers to the adjustment of the modeling factor (M_C) shown in Equation (3). This process was closely related to the fitting procedure described earlier. It was selected a M_C equal to 89.525×10^6 , which guaranteed that MCL levels did not exceed the maximum of 250 cu in any neural health condition used in the experiments. Stimulation above this limit would produce undesired out of compliance stimulation.

2.3. Neural health conditions

In preliminary experiments (not shown in this study) it was found that 9001 fibers introduced a considerable amount of

redundant information, and also, in the majority of the ANFs the node with the highest activation function for any electrode was beyond the fifth node. Therefore, it was decided to use different α_f for every ANF, but defining a mean degeneration value with a standard deviation of three nodes. This diminished the redundant information and was a more realistic representation of how the degeneration gradually occurs (Nadol, 1997).

The peripheral models with a mean ANF degeneration of 5, 10, and 15 nodes were considered to have a “healthy” neural health condition, a “moderate” loss, and a “severe” loss, respectively. Those three cases were assessed in this study and are shown in Figure 5.

2.4. Experiments

A total of nine SMT experiments and nine SRT experiments were performed using the three variants of the F120 sound coding strategies (F120-S, F120-P, and F120-T) and three aforementioned neural health conditions. Each neural health condition in the peripheral auditory model can be considered as an “individual” in experiments with real CI users.

2.4.1. Spectral modulation threshold experiments

The spectral modulation threshold (SMT) experiment was defined by Litvak et al. (2007b). It consists of measuring the smallest detectable spectral contrast in spectral rippled noise. The spectral shape of spectral rippled noise is sinusoidal and it is generated using Equation (9), where $|F(f_r)|$ is the magnitude at the frequency bin f_r , C_t is the spectral contrast in dB, f_{RPO} is the ripple-per-octave, and θ_0 is the ripple phase in the spectrum. Notice that the signal is hard-filtered at values below 350 Hz and above 5,600 Hz.

$$|F(f_r)| = \begin{cases} 10^{\frac{C_t}{20}} \sin(2\pi(\log_2(f_r/350))f_{RPO} + \theta_0)/20 & 350 < f_r < 5600 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

With human participants, the SMT is obtained with a three intervals two alternative forced choice procedure consisting of two reference intervals with no ripple (C_t equals to 0), and one target interval with a defined C_t . The first interval is always a reference noise, hence the participant has to indicate if the target interval is presented in the second or third position. This adaptive procedure is detailed by Litvak et al. (2007b), and has an equilibrium point of 79.4% correct answers. Because FADE uses a “training and testing” approach, the adaptive procedure could not be implemented; however, the equilibrium point is kept as the detection threshold.

The corpus was generated using Equation (9) with MATLAB. The ripple phase (θ_0) was randomly selected for every stimulus signal. A ripple per octave (f_{RPO}) equal to 0.5 was selected as in the experiments from Litvak et al. (2007b) and Langner et al. (2017).

The training corpus consisted of 1,000 samples of the spectral ripple noise with random integer values between 2 and 20 dB for contrast depth, and 1,000 samples of reference noise (C_t equals to 0). The testing corpus consisted of 10 sets, each one with 50 samples of reference noise and 50 samples of spectral ripple noise at a target C_t of 2, 3, 4, 5, 7, 9, 11, 14, 17, and 20 dB, respectively. In total 1,000 samples were predicted at 10 different contrast levels.

The sampling frequency of every sample was 17.4 kHz and the stimulus duration was limited to 0.4 s. In all cases, loudness roving was implemented keeping a mean value of -49 dB_{FS}, a roving peak of 5 dB_{FS}, and a roving resolution of 0.5 dB_{FS}.

The spectral modulation detection performance was described by the psychometric function in Equation (7), where p_{chance} was set to 50% because it was a binary decision. The SMT was the x value, where $P_s(x)$ was equal to 79.4%.

2.4.2. Speech reception threshold experiments

The SRT experiments were performed using the Oldenburg sentence test (OLSA). OLSA consists of a matrix sentence test of 50 words that belong to five different categories of 10 words each: name, verb, number, adjective, and noun. The sentences were constructed with one word from each category, following the previously mentioned order, giving a total of 10^5 possible combinations (Wagener et al., 1999). In a closed test procedure, the participants have previous knowledge of the words that may appear. Several sentences, mixed with noise at a specific SNR, are presented to the subject and the subject is asked to repeat them. A score based on the percentage of correctly recognized words is obtained. This is repeated at different SNR conditions and then a psychometric function is fitted to the obtained data points. The SNR value where this psychometric function crosses the 50% mark of correctly recognized words is the SRT result.

For SRT experiments, Schädler et al. (2016) and Jürgens et al. (2018) used a subset of 120 OLSA sentences to generate the training and testing corpus, but in this study, only a subset of 100 OLSA sentences was used to reduce computational resources. In this subset, each of the 50 words in the matrix appears exactly 10 times. A random excerpt of the noise provided by OLSA was added to the sentences at the required level to obtain the different SNRs, while the speech was kept at -49 dB_{FS}.

FADE uses a closed training/testing approach, meaning that the same sentences used in the training are used in the prediction stage. Therefore, the training corpus was generated with all the sentences in the subset mixed with noise at seven different SNR levels, from 0 to 18 dB in steps of 3 dB, and without noise, giving a total of 80 unique instances for each word. The testing corpus was generated with all the sentences mixed at 10 different SNR values, from -9 to 18 dB in steps of 3 dB, giving a total of 5,000 words to be predicted.

Regarding the psychometric function described in Equation (7), p_{chance} was set to 10% because it was a one word out of 10 decisions. The SRT was the x value where $P_s(x)$ was equal to 50%.

3. Results

3.1. Fitting

Figure 6 shows the MCL levels obtained for the computational model with the healthy, moderate degeneration, and severe degeneration condition. It also shows as a reference an ideal case (no degeneration in the ANFs), and the worst case (20 degenerated segments in the ANFs).

The MCL level difference across the 15 virtual channels between ideal and healthy conditions was only on average 2.87 CU.

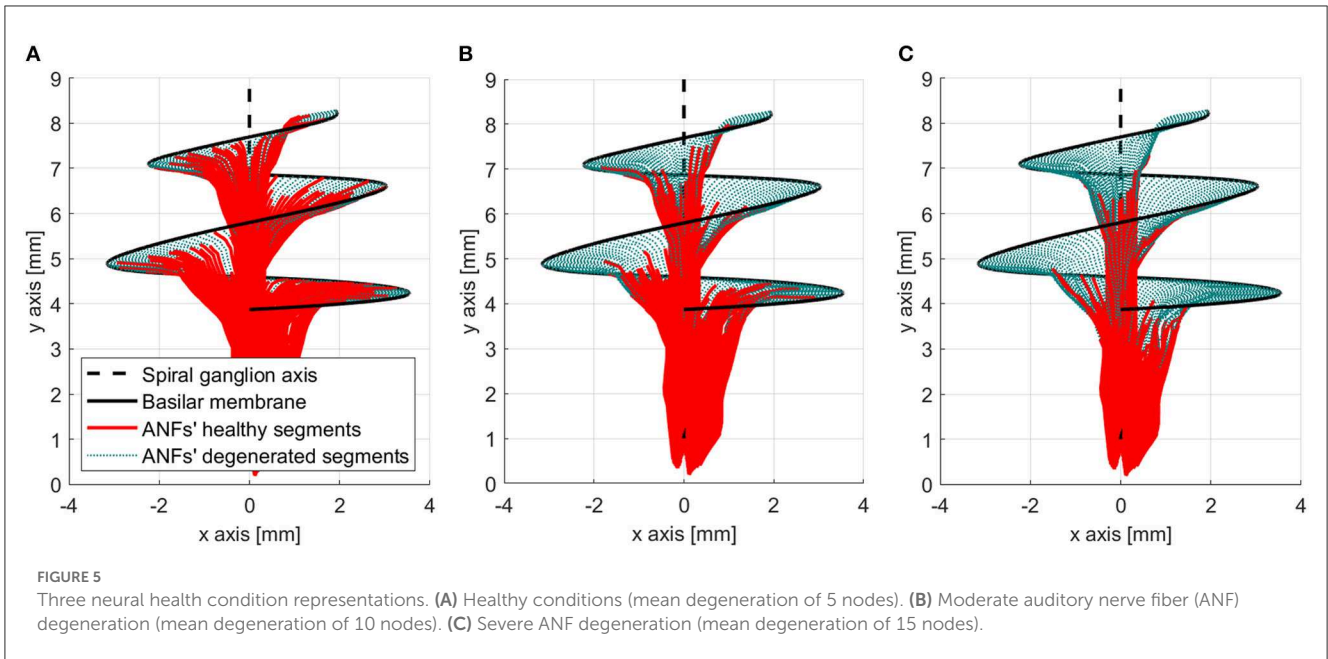


FIGURE 5 Three neural health condition representations. (A) Healthy conditions (mean degeneration of 5 nodes). (B) Moderate auditory nerve fiber (ANF) degeneration (mean degeneration of 10 nodes). (C) Severe ANF degeneration (mean degeneration of 15 nodes).

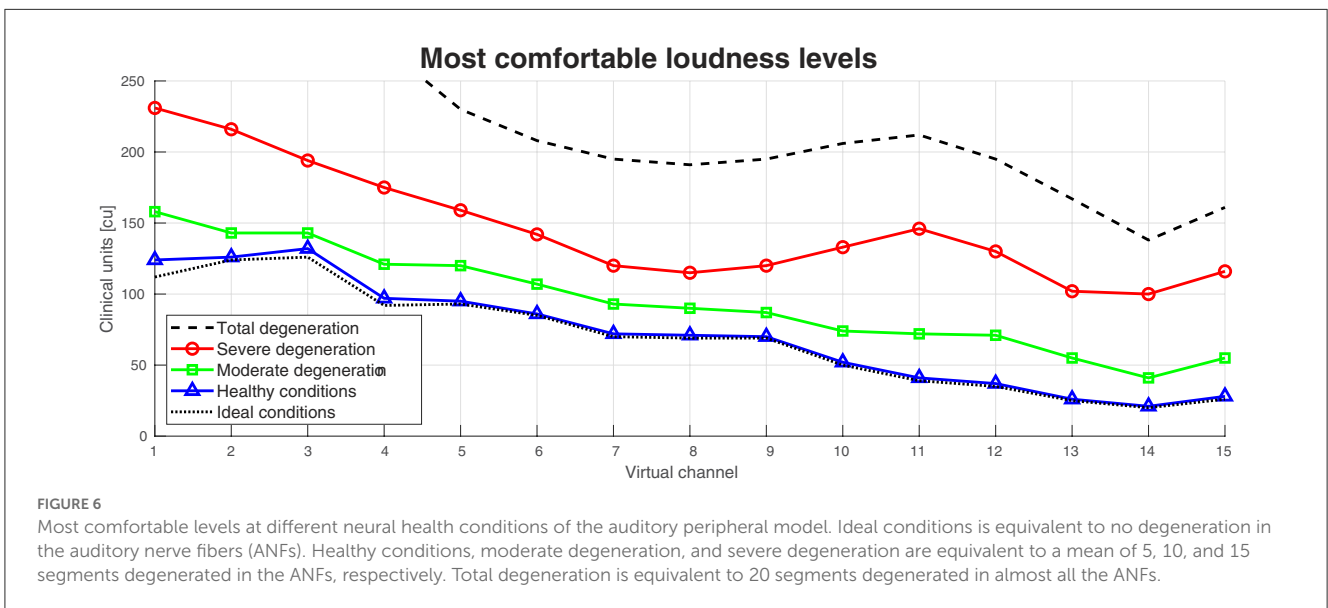


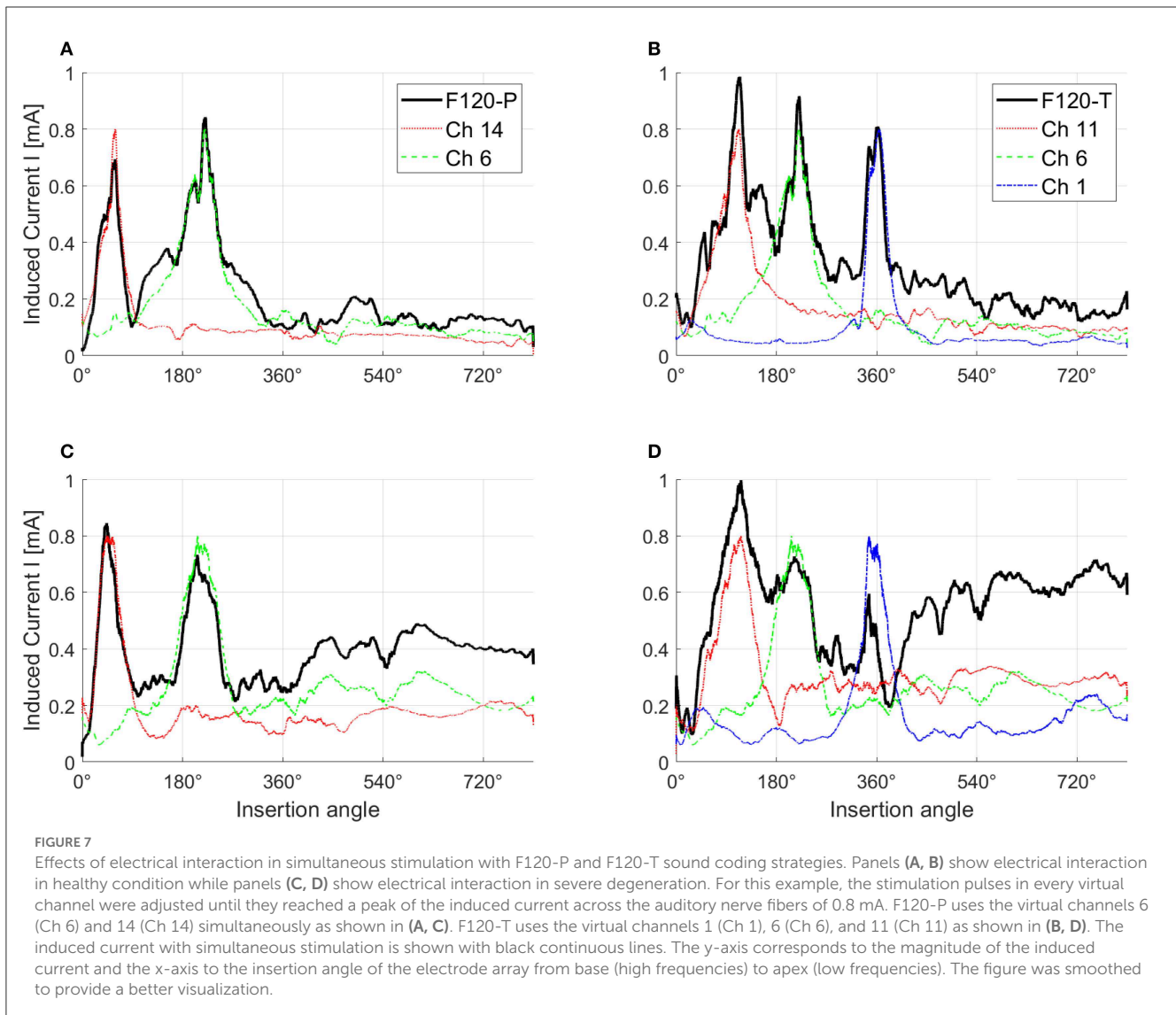
FIGURE 6 Most comfortable levels at different neural health conditions of the auditory peripheral model. Ideal conditions is equivalent to no degeneration in the auditory nerve fibers (ANFs). Healthy conditions, moderate degeneration, and severe degeneration are equivalent to a mean of 5, 10, and 15 segments degenerated in the ANFs, respectively. Total degeneration is equivalent to 20 segments degenerated in almost all the ANFs.

This difference increased with worse neural health conditions. Between healthy conditions and moderate degeneration, the difference was on average 23.47 CU, between moderate and severe degeneration, the difference was 51.27 CU, and between moderate and total degeneration, the difference was 80.53 CU. With total degeneration in the ANFs, the MCL levels of the four most basal electrodes (high frequencies) were above the desired 250 CU.

3.2. Electrical interaction

Figure 7 shows the effects of electrical interaction with simultaneous stimulation after fitting. For this figure, paired biphasic pulses were generated for the 1st, 6th, 11th, and

14th virtual channels to obtain a peak of induced current I across the ANF population of 0.8 mA. The black continuous lines correspond to the induced current with simultaneous stimulation, while the induced currents resulting from each channel individually (sequential stimulation) are shown with different colors. The induced current with simultaneous stimulation in healthy conditions (Figures 7A, B), and with paired stimulation in severe degeneration (Figure 7C), follows the peaks corresponding to the induced current of each individual virtual channel. This is not the case with triplet stimulation in severe degeneration (Figure 7D), where the peaks corresponding to virtual channels 6 and 14 are almost merged together while the peak corresponding to the virtual channel 1 is attenuated. Attenuation occurs when different virtual channels have opposite activation function signs; therefore, they cancel each other in Equation (3). Note



that the induced current toward the apex of the cochlea (insertion angles around 540° and 720°) also increases with worse neural health conditions as a consequence of higher stimulation levels.

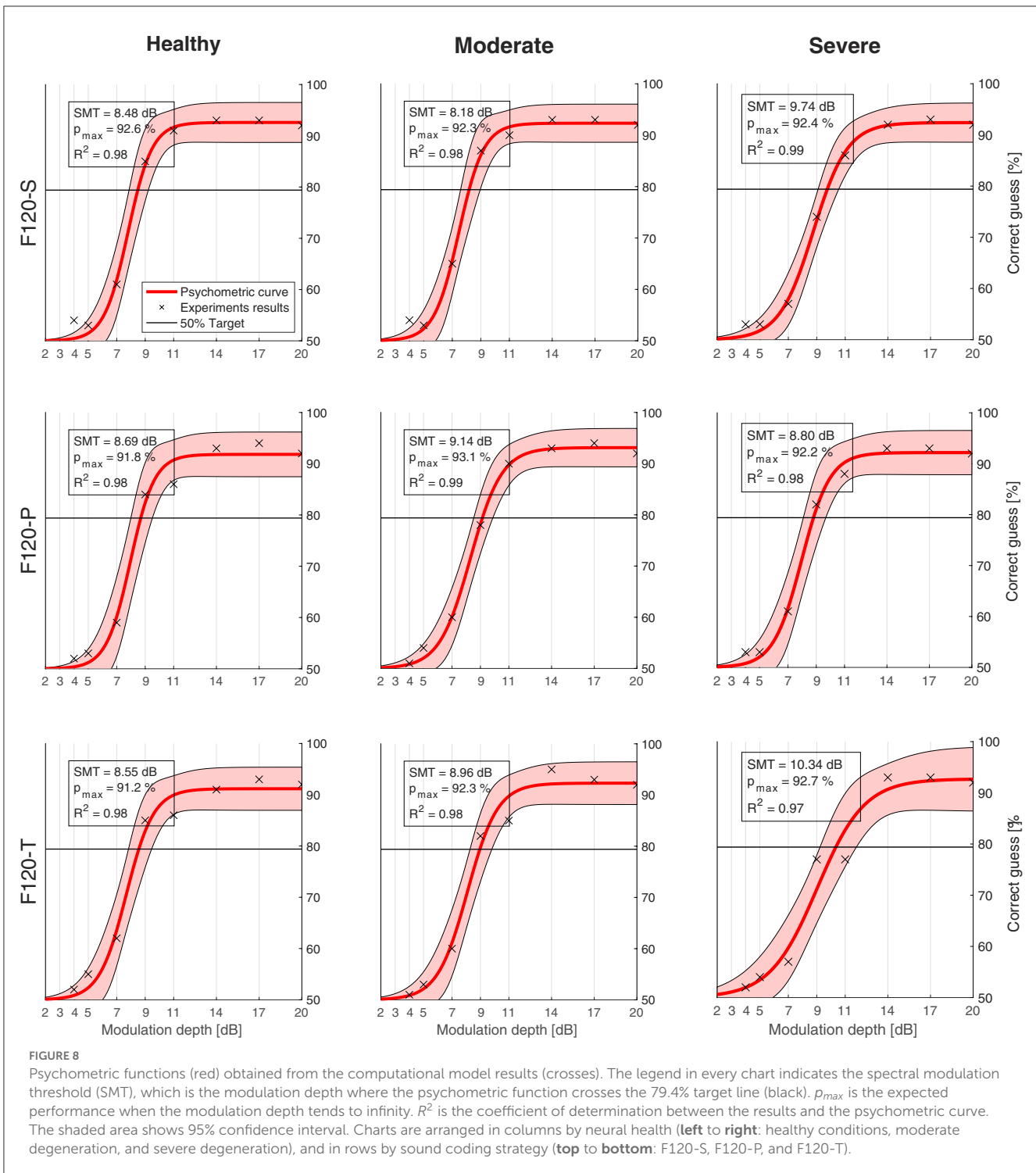
3.3. Spectral modulation threshold experiments

Figure 8 shows the psychometric functions obtained from the SMT experiments. The upper left box indicates the corresponding SMT, the expected performance with an infinite modulation depth (C_r), and the coefficient of determination (R^2). In general, the psychometric regression obtained an R^2 coefficient ranging from 0.97 to 0.99, which means that the psychometric function represents a good fit for the results obtained.

Nevertheless, Figure 9 shows that there was no significant effect on the performance regarding the sound coding strategy

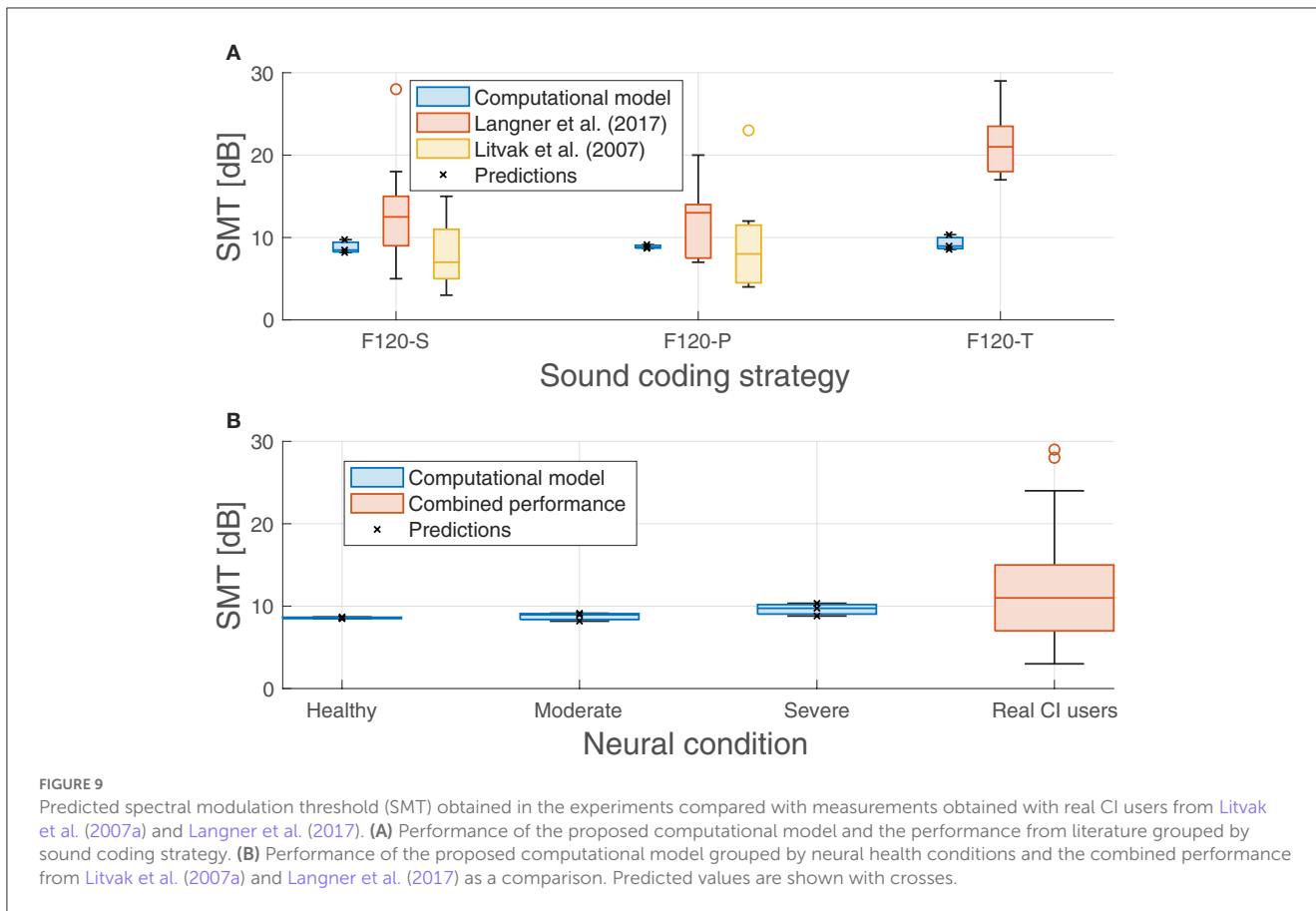
(Figure 9A) or the neural health condition (Figure 9B). As shown in Figure 9B, a small trend toward poorer performance with worse neural health conditions (8.57 dB for healthy conditions, 8.76 dB for moderate degeneration, and 9.63 dB for severe degeneration) is not significant compared to the clinical data obtained by Litvak et al. (2007a) and Langner et al. (2017). Also, as shown in Figure 9A, the expected effect of sound coding strategy on performance is not obtained (e.g., with severe degeneration the SMT is better using F120-P than using F120-S). The p_{max} seems to not be affected either by the neural degeneration or sound coding strategy since it varies from 91.2% (F120-T at healthy conditions) and 93.1% (F120-P at moderate degeneration of the ANFs).

Data from Litvak et al. (2007b) were collected from 25 CI users (Saoji et al., 2005) and are shown in Figure 9, part A, with blue boxes. The CI users were using F120-S (14 participants) and F120-P (11 participants) sound coding strategies. The results from Langner et al. (2017), shown in the same graph with green boxes, included 14



experiments. Half of them showed the SMT comparison between F120-S and F120-P sound coding strategies, while the other half was between F120-S and F120-T. Their performance showed a large variability but the trend is to get worse with simultaneous stimulation, especially with F120-T.

The performance of the computational model ranged from 8.48 to 10.34 dB, which can be considered as a “good” to “average” performance for CI users, but it does not account for the variability. The difference between the worst and the best performance of the model was only 1.90 dB, while in real CI users, it can be more than 18 dB.



3.4. Speech reception threshold experiments

Figure 10 shows the psychometric functions obtained from the SRT experiments. The upper left box in every chart indicates the corresponding SRT, the expected performance “in quiet” with p_{max} , and the coefficient of determination (R^2) obtained with the regression. In general, the psychometric function represented quite well the data resulted from the experiments, obtaining an R^2 coefficient always equal or greater than 0.99.

The effect of neural health is visible in the overall performance, affecting not only the SRT, but also the predicted performance in quiet. Figure 11 shows the SRT grouped by sound coding strategy (Figure 11A) and neural health condition (Figure 11B). In general, worse neural health conditions led to poorer performance (higher SRTs). The differences in SRT between healthy and severe degeneration conditions for F120-S, F120-P, and F120-T were 1.53, 1.63, and 2.77 dB_{SNR} , respectively. This indicates that simultaneous stimulation with F120-T was more sensitive to the effects of neural health degeneration than the F120-P.

On the contrary, the sound coding strategy that stimulated with two virtual channels simultaneously (F120-P) showed better overall performance, as shown in Figure 11, part A. With better neural health conditions, F120-T obtained better performance than F120-S, but with severe degeneration, the performance with F120-T felt below F120-S performance.

In addition, Figure 11 shows the results obtained by Jürgens et al. (2018). They measured the SRT in a group of 14 CI users using the advanced combinational encoder (ACE) sound coding strategy, which does not use current steering but it is comparable to F120-S because it uses sequential stimulation of biphasic pulses. In addition, they used a computational model to predict the performance of the CI users based on their individualized electrical field spread. Their measured SRTs ranged between -0.1 and 6.2 dB_{SNR} . In contrast, the performance of their computational model ranged between 4.67 and 7.56 dB_{SNR} , which is considered a poor performance according to Jürgens et al. (2018).

4. Discussion

In this study, a novel computational model to simulate the performance of CI users in psychoacoustic experiments was proposed. The proposed model consists of two main parts: (i) a front-end that includes a peripheral auditory model; (ii) a back-end based on the framework FADE. The peripheral auditory model combined a three-dimensional representation of the electrode-nerve interface taken from Nogueira et al. (2016), with a population of “phenomenologically” modeled ANFs taken from Joshi et al. (2017). This combination allowed to overcome the geometric limitations of the

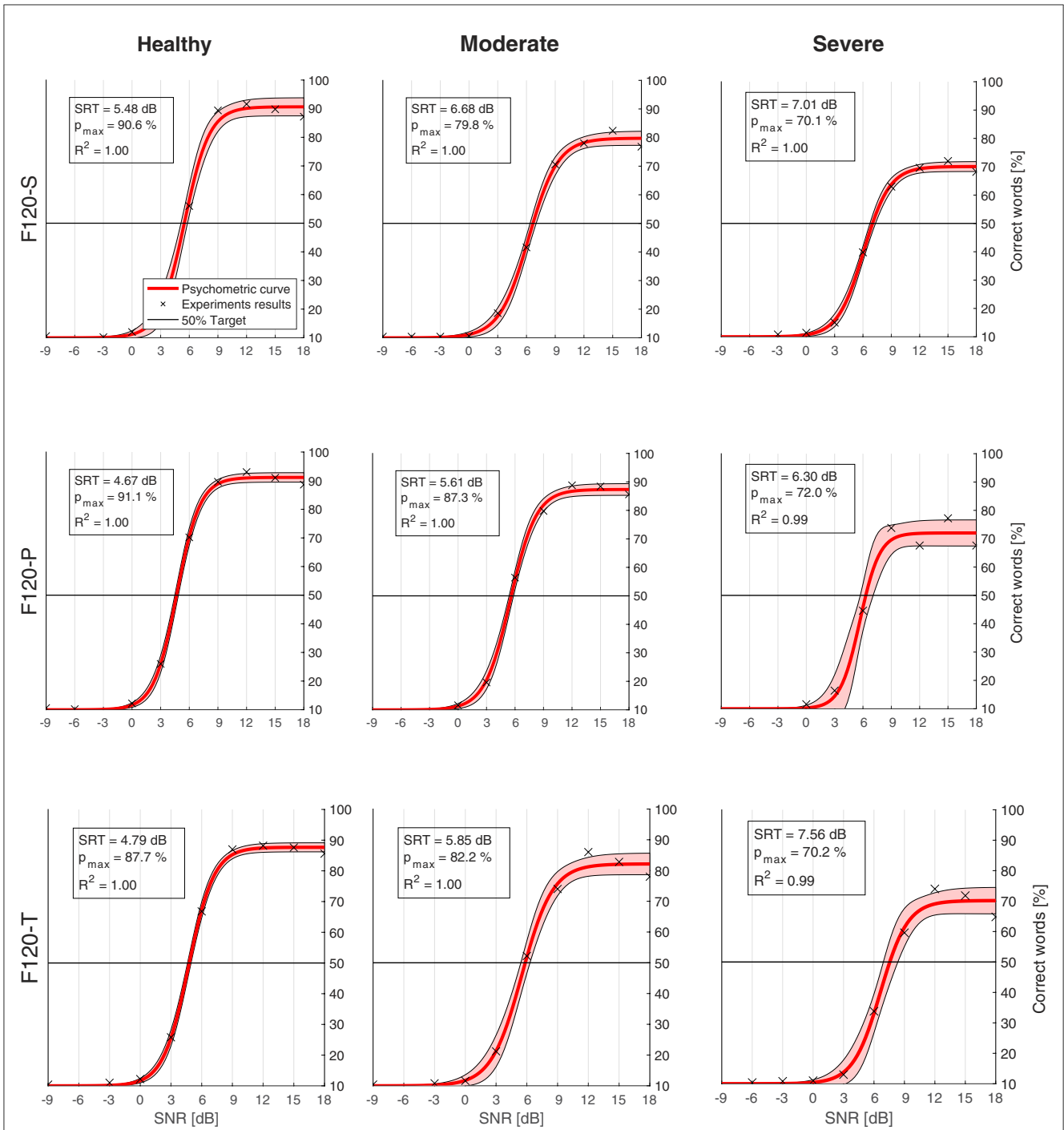


FIGURE 10 Psychometric functions (red) obtained from the computational model results (crosses). The legend in every chart indicates the speech reception threshold (SRT), which is the signal-to-noise ratio (SNR) where the psychometric function crosses the 50% target line (black). p_{max} is the expected performance in quiet. R^2 is the coefficient of determination between the results and the psychometric curve. The shaded area shows 95% confidence interval. Charts are arranged in columns by neural health (left to right: healthy conditions, moderate degeneration, and severe degeneration), and in rows by sound coding strategy (top to bottom: F120-S, F120-P, and F120-T).

phenomenological model by making the induced current dependent on the activation function, which was obtained using the “morphological” model of Ashida and Nogueira (2018). In that sense, the proposed peripheral auditory model takes advantage of the benefits of both “phenomenological” and “physiological” approaches.

This study assessed different neural health conditions by gradually removing segments of the ANFs from the periphery toward the spiral ganglion, which is an approximation to the real physiological degeneration (Nadol, 1988; Spoenlin and Schrott, 1988; Nogueira and Ashida, 2018). The activation function decreases and widens with worse neural health

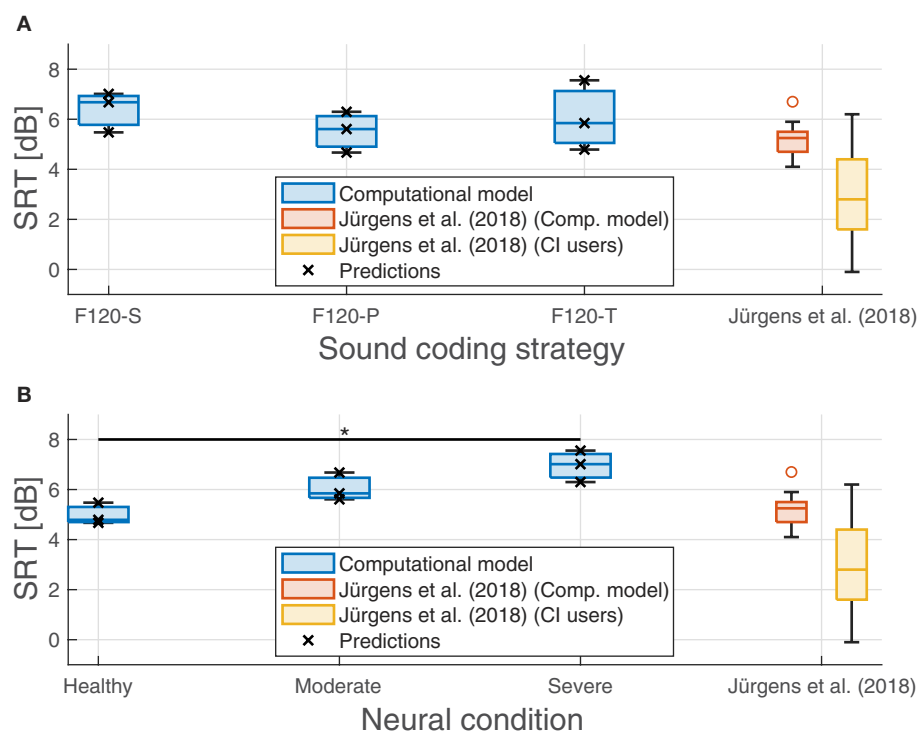


FIGURE 11

Predicted speech reception threshold (SRT) obtained in the experiments. **(A)** Performance of the proposed computational grouped by sound coding strategy. **(B)** Performance of the computational model grouped by neural health conditions. Predicted values are shown with crosses. Results from Jürgens et al. (2018) are shown as a reference in both panels as a reference. The red boxes show their predicted SRTs using an individualized model using the electrical field spread of real CI users and the yellow boxes show their measured SRTs. The symbol "*" denotes a statistical significance with a p -value less than 0.05 between healthy condition and severe degeneration.

conditions having an impact in the MCL levels, as shown in Figure 6. Worst neural health conditions resulted in higher MCL levels, which is consistent with the findings of Langner et al. (2021).

Predictions of the proposed computational model were obtained for two psychoacoustic experiments: SRT and SMT. The results of the SRT experiment show that worse neural health conditions result in poorer speech reception performance. Although this degeneration affected more the performance with simultaneous stimulation sound coding strategies, especially with F120-T, the detriment compared to sequential stimulation (F120-S) was rather small and not as relevant, as shown by Langner et al. (2017, 2021). A more detailed discussion regarding SMT and SRT experiments is presented in the following subsections.

4.1. Spectral modulation threshold experiments

As shown in Figure 9, the computational model performed similarly in SMT despite having different neural health conditions or sound coding strategies. This is because IR is a highly correlated feature along its dimensions (they are not orthogonal) and they cannot convey any relative spectral information

that can be used by the HMM to differentiate between spectral rippled noise and flat noise, especially when the phase θ_0 (see Equation 9) of the spectral rippled noise is randomized.

When the phase of the spectral ripple noise is randomized, the spectral peaks and valleys are always located in different auditory filters of the IR. With every realization of spectral ripple noise, the mean value of the IR gets closer to the mean value of the flat noise in every auditory filter. Only the standard deviation of the IR is always greater in spectral ripple noise than in flat noise. Therefore, whenever the neural activity in any auditory filter was greater, or lower, than the activity expected from the flat noise, the HMM classified it as spectral rippled noise. This may also explain why it consistently predicts an SMT around 9 dB since the loudness roving used to generate the training and testing corpus had a dynamic range of 10 dB.

Prediction algorithms based on HMMs work better with decorrelated features such as Mel frequency cepstral coefficients (MFCCs), Garbor filter bank (GFBs), or separable Garbor filter bank (SGFB) that are obtained from the spectral representation of the signal as shown by Kollmeier et al. (2016) and Schädler et al. (2016). On the contrary, IR is somehow equivalent to the bare spectrum of the audio signal, and it is highly correlation to work with HMMs.

4.2. Speech reception threshold experiments

Speech understanding relies on two principal aspects: temporal cues and spectral cues (Xu et al., 2005). Depending on the frequency, the temporal cues can be classified into envelopes (2–50 Hz), periodicity (50–500 Hz), and fine structure (500–10,000 Hz), the envelopes being specially important for speech understanding (Rosen, 1992). In the proposed computational model, the HMM was able to capture these envelopes along the auditory filters using Markov chains of eight states (Schädler et al., 2016); however, periodicity and fine structure is lost. The spectral cues, commonly related to the formants produced by the pronunciation of vowels and some consonants, were not properly captured by the HMM because of the before mentioned limitations of the IR when representing spectral shapes. Thus, words with relatively low SNR resulted in an IR much more similar to a word with a flat spectral shape. This may be the reason why, compared to real CI users, the performance of the computational model was worse than expected (higher SRTs) and did not account for the expected variability of around 6 dB as measured by Jürgens et al. (2018). Figure 11 also shows that this problem can be traced back to their computational model. Jürgens et al. (2018) also used the combination of IR as a feature and FADE as the back-end with an individualized electrical field spread that accounted for differences in the electrode–nerve interface between participants. However, the overall simulated performance resulted in higher SRTs than the one measured in real CI users as in our simulations.

In addition, there is a discrepancy between the effects of channel interaction in simultaneous stimulation (F120-P and F120-T) compared to sequential stimulation (F120-S). According to Langner et al. (2017, 2021), the performance with F120-S is similar to the performance with F120-P, but better than the performance with F120-T. This is not the case in the results shown in Figure 11, part A. This may be caused by the sharp decay of the voltage spread that is inversely proportional to the distance d_{nfa} (see Equation 1). A sharp decay diminishes the effect of electrical interaction between virtual channels in simultaneous stimulation; therefore, the performance may be more affected by the refractory period of ANFs. Notice in Figure 1 that with F120-S the electrical stimulation is continuous, while with F120-P and F120-T, there are stimulation gaps that may help ANFs to recover from refractoriness.

However, Figure 7 shows that the presented computational model is capable of reproducing the effects of electrical interaction between virtual channels. This effect is larger with the F120-T sound coding strategy because the virtual channels are closer together. Electrical interaction obtained with F120-P is much lower compared to F120-T. This may explain why in real CI users the performance with F120-P is similar to the performance with F120-S, but considerably worse with F120-T (Langner et al., 2017, 2020b). Figures 7, 11, part B, also support the hypothesis assessed in this study since the performance with triplet stimulation was significantly affected by severe degeneration compared to healthy conditions.

4.3. Future improvements

4.3.1. Feature extraction and back-end

As discussed earlier, the IR proposed by Fredelake and Hohmann (2012) may not be the best set of features to use with the HMM already implemented in FADE because IR is highly correlated within its own dimensions. Therefore, it does not carry any spectral shape information that indicates the relative neural activity between the auditory filters. A solution could be to incorporate other decorrelated features that have been shown to improve the performance of automatic speech recognition (ASR) algorithms and that also worked with neural activity (Holmberg et al., 2005, 2007; Nogueira et al., 2007). But, although these algorithms may provide benefits in performance for the computational model, these may not represent any particular physiological aspect of the auditory system, which was the idea behind the proposed computational model.

The central processes that occur in the auditory pathway beyond the peripheral auditory system are not completely understood. The IR is a simple model that makes many assumptions about how spikes are processed centrally to interpret sounds. However, the IR was necessary to accommodate the neural activity to the temporal resolution and number of features adequate for an HMM back-end. Another way to approach this problem would be to substitute FADE as the back-end for an algorithm that can perform predictions directly from the spike activity coming from the peripheral auditory model (Alvarez and Nogueira, 2022), but this is a challenging task due to the amount of ANFs modeled and the sample rate of the spike activity. Artificial neural networks (ANNs) seem to suit well with this approach (Kell et al., 2018; Santana et al., 2018; Wang et al., 2018) since ANNs can manage a large amount of data and there is no intrinsic assumption of any central auditory process. Neural networks could function as a “black box” while the detailed modeling is focused on the peripheral auditory system (Brochier et al., 2022).

4.3.2. Peripheral auditory model

Although the proposed peripheral auditory model can account for many physiological aspects, there is room for improvement.

The voltage spread obtained in this study presents a decay inversely proportional to the distance. This results in a sharper decay compared to the exponential decay measured in laboratories with saline solutions (O’leary et al., 1985; Kral et al., 1998). The sharper the decay, the less electrical interaction between channels in simultaneous stimulation; therefore, the model is less sensitive to the differences between the F120 variants studied. Nevertheless, a more realistic voltage spread could be obtained by using finite element method (FEM) or a boundary element method (BEM) in the three-dimensional electrode–nerve instead of assuming a homogeneous medium (Kalkman et al., 2014, 2022; Nogueira et al., 2016; Croner et al., 2022). This improvement of the peripheral auditory model would take into account the electrical characteristics of the bones, tissue, and other media present in the cochlea.

Regarding the ANF model, the morphological model used in this study considers the ANF as a cable with homogeneously

distributed segments that do not differentiate between peripheral axon, central axon, or the soma. This approach worked because it allowed us to adjust the induced current I for the physiological model of Joshi et al. (2017) depending on different aspects such as electrode and nerve position, the direction of the ANF with respect to the electrode, and the effects of degeneration. A more realistic multi-compartment cable model that represents its morphology and physiology more accurately (Rattay et al., 2013; Bachmaier et al., 2019; Kalkman et al., 2022) could be implemented; however, one must be careful when coupling it with the model of Joshi et al. (2017). If the morphology already takes into account differences between the axons and the soma, the original parameters used in Equation (5) are not validated anymore.

Another improvement could be a new degeneration method that takes into account, not only the progressive inhibition of nodes, but the diameter reduction in the axons (Heshmat et al., 2020; Croner et al., 2022). In the current degeneration method, the number of degenerated nodes is not equal across the ANF population, instead, the number of nodes degenerated in each ANF is governed by a mean value and a standard deviation that is arbitrarily set to three nodes. This is an assumption made since there is no available information on how it is in real implanted cochleas. In fact, it is probable that the fibers degenerate in different patterns, for example, more degeneration in the basal turn than in the apical turns. Therefore, in further studies related to degeneration, different patterns should be taken into account including dead regions in the cochlea, which are relevant in electrical stimulation (Moore et al., 2010).

5. Conclusion

The computational model presented in this study was capable of executing simulations of SRT and SMT experiments. It consisted of a peripheral auditory model with a three-dimensional electrode-nerve interface that allows to represent different neural health conditions by applying a systematic degeneration to the modeled ANFs. The neural health condition affected the fitting procedure and speech reception in the expected manner, augmenting the current needed to reach MCLs and obtaining worst (higher) SRTs, respectively.

The computational model could not reproduce quantitatively the expected results in SRT experiments from real CI users where simultaneous stimulation sound coding strategies (F120-P and F120-T) consistently performed worse than sequential stimulation sound coding strategies (F120-S). Nevertheless, the results showed that the qualitative performance detriment due to neural health conditions with simultaneous stimulation (especially with F120-T) was higher than with sequential stimulation.

SMT experiments with the computational model were inconclusive since the results showed no relevant impact neither from neural health conditions nor channel interaction caused by the simultaneous stimulation. This is arguably caused by the selected IR resulting in features that did not convey spectral shape information, together with an HMM-based recognizer.

Future developments of the computational model could offer a reliable tool to assess the effects of different sound coding

strategies and different neural health conditions in psychoacoustic experiments without the need for testing in implanted volunteers, especially, in the early development stages of new CI technologies. The improvements should be focused on the physiological model of the ANFs and the feature extraction from the neural activity.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

FA designed the experiments, programmed and integrated the computational model, verified and analyzed the results, and wrote the original manuscript draft. DK developed the auditory nerve fiber model. WN supervised the research project. All authors contributed to conceptualization, contributed to manuscript revision, read, and approved the submitted version.

Funding

This project was funded by the German Academic Exchange Service (DAAD) under the Research Grants – Doctoral Programmes in Germany 2021/2022 program, and was co-funded by the Cluster for Excellence Hearing4All. This work is part of the project that received funding from the European Research Council (ERC) under the European Union's Horizon-ERC Program (Grant agreement READIHEAR No. 101044753-PI: WN).

Acknowledgments

Special thanks to Dr. Florian Langner for his help in the early stage of the project, specially, for providing a better insight of the sound coding strategies used in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Advanced Bionics (2020). *Target CI Fitting Guide*.
- Alvarez, F., and Nogueira, W. (2022). "Predicting speech intelligibility using the spike acativity mutual information index," in *Proc. Interspeech 2022* (Incheon), 3503–3507.
- Ashida, G., and Nogueira, W. (2018). Spike-conducting integrate-and-fire model. *eNeuro* 5. doi: 10.1523/ENEURO.0112-18.2018
- Bachmaier, R., Encke, J., Obando-Leiton, M., Hemmert, W., and Bai, S. (2019). Comparison of multi-compartment cable models of human auditory nerve fibers. *Front. Neurosci.* 13, 1173. doi: 10.3389/fnins.2019.01173
- Brochier, T., Schlittenlacher, J., Roberts, I., Goehring, T., Jiang, C., Vickers, D., et al. (2022). From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants. *IEEE Trans. Biomed. Eng.* 69, 3300–3312. doi: 10.1109/TBME.2022.3167113
- Bruce, L. C., White, M. W., Irlight, L. S., O'Leary, S. J., Dynes, S., Javel, E., et al. (1999). A stochastic model of the electrically stimulated auditory nerve: single-pulse response. *IEEE Trans. Biomed. Eng.* 46, 617–629. doi: 10.1109/10.764938
- Chen, F., and Loizou, P. C. (2011). Predicting the intelligibility of vocoded speech. *Ear Hear.* 32, 331–338. doi: 10.1097/AUD.0b013e3181ff3515
- Croner, A. M., Heshmat, A., Schrott-Fischer, A., Glueckert, R., Hemmert, W., and Bai, S. (2022). Effects of degrees of degeneration on the electrical excitation of human spiral ganglion neurons based on a high-resolution computer model. *Front. Neurosci.* 16, 914876. doi: 10.3389/fnins.2022.914876
- Dillon, M. T., Buss, E., King, E. R., Deres, E. J., Obarowski, S. N., Anderson, M. L., et al. (2016). Comparison of two cochlear implant coding strategies on speech perception. *Cochlear Implants Int.* 17, 263–270. doi: 10.1080/14670100.2016.1244033
- Dorman, M. F., Loizou, P. C., and Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J. Acoust. Soc. Am.* 102, 2993–2996. doi: 10.1121/1.420354
- El Boghdady, N., Kegel, A., Lai, W. K., and Dillier, N. (2016). A neural-based vocoder implementation for evaluating cochlear implant coding strategies. *Hear. Res.* 333, 136–149. doi: 10.1016/j.heares.2016.01.005
- Fredelake, S., and Hohmann, V. (2012). Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation. *Hear. Res.* 287, 76–90. doi: 10.1016/j.heares.2012.03.005
- Gajecki, T., and Nogueira, W. (2021). "An end-to-end deep learning speech coding and denoising strategy for cochlear implants," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Singapore).
- Hamacher, V. (2004). *Signalverarbeitungsmodelle des elektrisch stimulierten Gehors* (Ph.D. thesis). RWTH Aachen, Wissenschaftsverlag Mainz, Aachen, Germany.
- Han, J. H., Lee, H. J., Kang, H., Oh, S. H., and Lee, D. S. (2019). Brain plasticity can predict the cochlear implant outcome in adult-onset deafness. *Front. Hum. Neurosci.* 13, 38. doi: 10.3389/fnhum.2019.00038
- Heshmat, A., Sajedi, S., Chacko, L. J., Fischer, N., Schrott-Fischer, A., and Rattay, F. (2020). Dendritic degeneration of human auditory nerve fibers and its impact on the spiking pattern under regular conditions and during cochlear implant stimulation. *Front. Neurosci.* 14, 599868. doi: 10.3389/fnins.2020.599868
- Heshmat, A., Sajedi, S., Schrott-Fischer, A., and Rattay, F. (2021). Polarity sensitivity of human auditory nerve fibers based on pulse shape, cochlear implant stimulation strategy and array. *Front. Neurosci.* 15, 751599. doi: 10.3389/fnins.2021.751599
- Hochmair-Desoyer, I., Schulz, E., Moser, L., and Schmidt, M. (1997). The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. *Am. J. Otol.* 18, S83–S86.
- Holmberg, M., Gelbart, D., and Hemmert, W. (2007). Speech encoding in a model of peripheral auditory processing: quantitative assessment by means of automatic speech recognition. *Speech Commun.* 49, 917–932. doi: 10.1016/j.specom.2007.05.009
- Holmberg, M., Gelbart, D., Ramacher, U., and Hemmert, W. (2005). "Automatic speech recognition with neural spike trains," in *9th European Conference on Speech Communication and Technology* (Lisbon), 1253–1256.
- ImSiecke, M., Büchner, A., Lenarz, T., and Nogueira, W. (2021). Amplitude growth functions of auditory nerve responses to electric pulse stimulation with varied interphase gaps in cochlear implant users with ipsilateral residual hearing. *Trends Hear.* 25. doi: 10.1177/23312165211014137
- Joshi, S. N., Dau, T., and Epp, B. (2017). A model of electrically stimulated auditory nerve fiber responses with peripheral and central sites of spike generation. *J. Assoc. Res. Otolaryngol.* 18, 323–342. doi: 10.1007/s10162-016-0608-2
- Jürgens, T., Hohmann, V., Büchner, A., and Nogueira, W. (2018). The effects of electrical field spatial spread and some cognitive factors on speech-in-noise performance of individual cochlear implant users—A computer model study. *PLoS ONE* 13, e0193842. doi: 10.1371/journal.pone.0193842
- Kalkman, R. K., Briaire, J. J., Dekker, D. M., and Frijns, J. H. (2014). Place pitch versus electrode location in a realistic computational model of the implanted human cochlea. *Hear. Res.* 315, 10–24. doi: 10.1016/j.heares.2014.06.003
- Kalkman, R. K., Briaire, J. J., Dekker, D. M., and Frijns, J. H. (2022). The relation between polarity sensitivity and neural degeneration in a computational model of cochlear implant stimulation. *Hear. Res.* 415, 108413. doi: 10.1016/j.heares.2021.108413
- Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 98, 630.e16–644.e16. doi: 10.1016/j.neuron.2018.03.044
- Kollmeier, B., Schädler, M. R., Warzybok, A., Meyer, B. T., and Brand, T. (2016). Sentence recognition prediction for hearing-impaired listeners in stationary and fluctuation noise with FADE. *Trends Hear.* 20, 1–17. doi: 10.1177/2331216516655795
- Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (1998). Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents. *Hear. Res.* 121, 11–28. doi: 10.1016/S0378-5955(98)00061-6
- Landsberger, D. M., and Srinivasan, A. G. (2009). Virtual channel discrimination is improved by current focusing in cochlear implant recipients. *Hear. Res.* 254, 34–41. doi: 10.1016/j.heares.2009.04.007
- Langner, F., Arenberg, J. G., Buchner, A., and Nogueira, W. (2021). Assessing the relationship between neural health measures and speech performance with simultaneous electric stimulation in cochlear implant listeners. *PLoS ONE* 16, e0261295. doi: 10.1371/journal.pone.0261295
- Langner, F., Buchner, A., and Nogueira, W. (2020a). Evaluation of an adaptive dynamic compensation system in cochlear implant listeners. *Trends Hear.* 24. doi: 10.1177/2331216520970349
- Langner, F., McKay, C. M., Büchner, A., and Nogueira, W. (2020b). Perception and prediction of loudness in sound coding strategies using simultaneous electric stimulation. *Hear. Res.* 398. doi: 10.1016/j.heares.2020.108091
- Langner, F., Saoji, A. A., Büchner, A., and Nogueira, W. (2017). Adding simultaneous stimulating channels to reduce power consumption in cochlear implants. *Hear. Res.* 345, 96–107. doi: 10.1016/j.heares.2017.01.010
- Litvak, L. M., Spahr, A. J., and Emadi, G. (2007a). Loudness growth observed under partially tripolar stimulation: model and data from cochlear implant listeners. *J. Acoust. Soc. Am.* 122, 967–981. doi: 10.1121/1.2749414
- Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (2007b). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J. Acoust. Soc. Am.* 122, 982–991. doi: 10.1121/1.2749413
- Malherbe, T. K., Hanekom, T., and Hanekom, J. J. (2016). Constructing a three-dimensional electrical model of a living cochlear implant user's cochlea. *Int. J. Num. Methods Biomed. Eng.* 32. doi: 10.1002/cnm.2751
- McKay, C. M., and McDermott, H. J. (1998). Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals. *J. Acoust. Soc. Am.* 104, 1061–1074.
- McKay, C. M., Remine, M. D., and McDermott, H. J. (2001). Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation. *J. Acoust. Soc. Am.* 110, 1514–1524. doi: 10.1121/1.1394222
- Moberly, A. C., Lowenstein, J. H., and Nittrouer, S. (2016). Word recognition variability with cochlear implants: "perceptual attention" versus "auditory sensitivity". *Ear Hear.* 37, 14–26. doi: 10.1097/AUD.0000000000000204
- Moore, B. C. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otol. Neurotol.* 24, 243–254. doi: 10.1097/00129492-200303000-00019
- Moore, B. C. J., Glasberg, B., and Schlueter, A. (2010). Detection of dead regions in the cochlea: relevance for combined electric and acoustic stimulation. *Adv. Otorhinolaryngol.* 67, 43–50. doi: 10.1159/000262595
- Nadol, J. (1997). Patterns of neural degeneration in the human cochlea and auditory nerve: implications for cochlear implantation. *Otolaryngol. Head Neck Surg.* 117(3 Pt 1), 220–228. doi: 10.1016/s0194-5998(97)70178-5
- Nadol, J. B. (1988). Comparative anatomy of the cochlea and auditory nerve in mammals. *Hear. Res.* 34, 253–266.
- Nelson, P. B., and Jin, S.-H. (2004). Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. *J. Acoust. Soc. Am.* 115, 2286–2294. doi: 10.1121/1.1703538
- Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (2003). Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J. Acoust. Soc. Am.* 113, 961–968. doi: 10.1121/1.1531983
- Nogueira, W., and Ashida, G. (2018). *Development of a Parametric Model of the Electrically Stimulated Auditory Nerve*. Cham: Springer International Publishing.

- Nogueira, W., Büchner, A., Lenarz, T., and Edler, B. (2005). A psychoacoustic “NofM”-type speech coding strategy for cochlear implants. *Eurasip J. Appl. Signal Process.* 2005, 3044–3059. doi: 10.1155/ASP.2005.3044
- Nogueira, W., El Boghdady, N., Langner, F., Gaudrain, E., and Bakent, D. (2021). Effect of channel interaction on vocal cue perception in cochlear implant users. *Trends Hear.* 25. doi: 10.1177/23312165211030166
- Nogueira, W., Harczos, T., Edler, B., Ostermann, J., and Büchner, A. (2007). “Automatic speech recognition with a cochlear implant front-end,” in *International Speech Communication Association - 8th Annual Conference of the International Speech Communication Association, Interspeech 2007* (Antwerp), 2537–2540.
- Nogueira, W., Litvak, L., Edler, B., Ostermann, J., and Büchner, A. (2009). Signal processing strategies for cochlear implants using current steering. *Eurasip J. Adv. Signal Process.* 2009. doi: 10.1155/2009/531213
- Nogueira, W., Schurzig, D., Büchner, A., Penninger, R. T., and Würfel, W. (2016). Validation of a cochlear implant patient-specific model of the voltage distribution in a clinical setting. *Front. Bioeng. Biotechnol.* 4, 84. doi: 10.3389/fbioe.2016.00084
- O’leary, S. J., Black, R. C., and Clark, G. M. (1985). Current distributions in the cat cochlea: a modelling and electrophysiological study. *Hear. Res.* 18, 273–281. doi: 10.1016/0378-5955(85)90044-9
- Prado-Gutierrez, P., Fewster, L. M., Heasman, J. M., McKay, C. M., and Shepherd, R. K. (2006). Effect of interphase gap and pulse duration on electrically evoked potentials is correlated with auditory nerve survival. *Hear. Res.* 215, 47–55. doi: 10.1016/j.heares.2006.03.006
- Ramekers, D., Versnel, H., Strahl, S. B., Smeets, E. M., Klis, S. F. L., and Grolman, W. (2014). Auditory-nerve responses to varied inter-phase gap and phase duration of the electric pulse stimulus as predictors for neuronal degeneration. *J. Assoc. Res. Otolaryngol.* 15, 187–202. doi: 10.1007/s10162-013-0440-x
- Rattay, F. (1999). The basic mechanism for the electrical stimulation of the nervous system. *Neuroscience* 89, 335–346.
- Rattay, F., Leao, R. N., and Felix, H. (2001). A model of the electrically excited human cochlear neuron. II. influence of the three-dimensional cochlear structure on neural excitability. *Hear. Res.* 153, 64–79. doi: 10.1016/S0378-5955(00)00257-4
- Rattay, F., Potrusil, T., Wenger, C., Wise, A. K., Glueckert, R., and Schrott-Fischer, A. (2013). Impact of morphometry, myelination and synaptic current strength on spike conduction in human and cat spiral ganglion neurons. *PLoS ONE* 8, e79256. doi: 10.1371/journal.pone.0079256
- Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. Biol. Sci.* 336, 367–373.
- Santana, L. M. Q. D., Santos, R. M., Matos, L. N., and Macedo, H. T. (2018). Deep neural networks for acoustic modeling in the presence of noise. *IEEE Latin Am. Trans.* 16, 918–925. doi: 10.1109/TLA.2018.8358674
- Santos, J. F., Consentino, S., Hazrati, O., Loizou, P. C., and Falk, T. H. (2013). Objective speech intelligibility measurement for cochlear implant users in complex listening environments. *Speech Commun.* 55, 815–824. doi: 10.1016/j.specom.2013.04.001
- Saoji, A. A., Litvak, L., Emadi, G., and Spahr, A. J. (2005). “Spectral modulation transfer function in cochlear implant listeners,” in *Conference on Implantable Auditory Prostheses* (Asilomar, CA).
- Schädler, M. R., Warzybok, A., Ewert, S. D., and Kollmeier, B. (2016). A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. *J. Acoust. Soc. Am.* 139, 2708–2722. doi: 10.1121/1.4948772
- Schädler, M. R., Warzybok, A., and Kollmeier, B. (2018). Objective prediction of hearing aid benefit across listener groups using machine learning: speech recognition performance with binaural noise-reduction algorithms. *Trends Hear.* 22, 1–21. doi: 10.1177/2331216518768954
- Smit, J. E., Hanekom, T., and Hanekom, J. J. (2008). Predicting action potential characteristics of human auditory nerve fibres through modification of the Hodgkin-Huxley equations. *South Afr. J. Sci.* 104, 284–292.
- Spoendlin, H., and Schrott, A. (1988). The spiral ganglion and the innervation of the human organ of corti. *Acta Otolaryngol.* 105, 403–410.
- Stadler, S., and Leijon, A. (2009). Prediction of speech recognition in cochlear implant users by adapting auditory models to psychophysical data. *Eurasip J. Adv. Signal Process.* 2009. doi: 10.1155/2009/175243
- Wagener, K., Kuehnel, V., and Kollmeier, B. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache I: design des oldenburger satztests. *Zeitschrift für Audiologie* 38, 4–15.
- Wang, W., Pedretti, G., Milo, V., Carboni, R., Calderoni, A., Ramaswamy, N., et al. (2018). Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses. *Sci. Adv.* 4, 4752–4764. doi: 10.1126/sciadv.aat4752
- Wouters, J., McDermott, H. J., and Francart, T. (2015). Sound coding in cochlear implants: from electric pulses to hearing. *IEEE Signal Process. Mag.* 32, 67–80. doi: 10.1109/MSP.2014.2371671
- Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405