DOI: 10.61782/fa.2023.0883

forum acusticum 2023

MACHINE-LEARNING-BASED AUDIO ALGORITHMS FOR HEARING LOSS COMPENSATION

Marjoleen Wouters[1*] **Fotios Drakopoulos**[1] **Sarah Verhulst**[1]
[1] Hearing Technology lab @WAVES, Department of Information Technology, Ghent University, Ghent, Belgium

ABSTRACT

Computational auditory models have been used for decades to develop audio signal processing algorithms in hearing aids (HAs). Here, using a biophysically inspired auditory model in a differentiable convolutional-neural-network (CNN) description (CoNNear), we trained end-to-end machine-learning- (ML) based audio signal-processing algorithms that maximally restored auditory-nerve (AN) responses affected by cochlear synaptopathy. To this end, we used backpropagation to develop several ML-based algorithms that match the simulated response of the corresponding hearing-impaired model back to the normal-hearing response, each time using the same CNN encoder-decoder architecture but different loss functions to achieve different compensation of the AN responses. Evaluation of the HA models was performed by processing sentences of the Flemish matrix test and comparing model outcomes with the unprocessed sentences. The magnitude spectra of all processed sentences showed differences between the HA models in amplification of low- and high-frequency speech content, whereas the high-frequency processing often introduced audible tonal distortions. Our processing showed different enhancement of the AN population responses at the speech onsets, vowels and consonants. We will objectively assess the effect of the most optimal compensation algorithms on sound quality and speech intelligibility in future clinical experiments.

Keywords: *hearing-aid processing, machine-learning, cochlear synaptopathy.*

1. INTRODUCTION

Exposure to noise or ototoxic drugs and aging are common causes of sensorineural hearing loss (SNHL) in humans, and often result in irreversible damage of the outer hair cells (OHCs) or synapses to the auditory nerve (AN), i.e. cochlear synaptopathy (CS) [1-3]. Several studies have suggested that CS results in a loss of the low- (LSR), medium- (MSR) and high-spontaneous rate (HSR) AN fibers (ANFs), in which the LSR and MSR are the first to be lost [2]. CS degrades encoding of the temporal envelope in sound, which may contribute to a variety of perceptual abnormalities such as speech-in-noise difficulties and decreased speech intelligibility [2, 4, 5].

However, pure tone audiometric thresholds, related to OHC loss, are not affected in CS, therefore CS is referred to as "hidden hearing loss" [6]. Studies on animal models have shown that the loss of ANFs and synapses to the AN, related to cochlear neuropathy and synaptopathy, are the first signs of permanent hearing damage and occur earlier in time than OHC loss [1,7]. Since the audiogram is an insensitive marker for damage to the AN and loss of synapses, patients suffering from CS will experience difficulties understanding speech in challenging situations while their hearing thresholds remain normal. Thus, it is expected that a large group of the noise-exposed or aging population suffers from CS, which still remains undiagnosed based on their audiogram and will

[*]***Corresponding author***: marjoleen.wouters@ugent.be

therefore not be treated properly. Non-invasive diagnostic techniques of CS have been recently proposed based on auditory-evoked potentials (AEPs) [8].

The current hearing-aid (HA) algorithms focus on compensating for the elevated audiometric thresholds, but do not specifically compensate for the hearing difficulties related to CS, and therefore offer no treatment to patients who suffer from CS. The non-linear dynamic-range compression of current HAs reduces the amplitude fluctuations of the temporal envelope, which might even worsen the hearing ability in case of CS [9-11]. HA algorithms aiming to compensate for OHC loss and CS-related hearing impairment hence need to be fundamentally different from standard HA algorithms, and be able to grasp the complex non-linear working mechanism of the auditory system.

Auditory models have been used for decades to develop audio signal processing algorithms in HAs. Typically, the difference signal between a normal-hearing (NH) and hearing-impaired (HI) model is used to design such algorithms, but only recently machine-learning (ML) methods have made their entry in this field. Specifically, when adopting differentiable descriptions of biophysical models of hearing impairment, it is possible to fully backpropagate through the models and design a new type of ML-based audio signal processing that compensates for different aspects of SNHL [12-13]. The objective of this work is to investigate several ML-based HA algorithms able to restore CS, based on a convolutional neural network (CNN) description of a NH and HI auditory model. We will also investigate which sound and speech features are modified when letting these ML algorithms decide the most-optimal solution to compensate for CS.

2. METHODS

2.1 CoNNear Auditory Model

We used a convolutional neural network model of the auditory periphery, CoNNear [14-16], that was developed starting from a biophysically inspired computational model of the human auditory periphery [17]. The CoNNear model provides a fast and differentiable description of the stages (basilar membrane (BM) vibrations, inner-hair-cell (IHC) potential, AN firing) of the human auditory system across 201 simulated tonotopic cochlear locations, with center frequencies (CFs) spaced according to the Greenwood place-frequency map of the cochlea [18].

The NH CoNNear model is shown in Fig. 1, and can simulate the AN response r_F to an auditory input x, sampled at a rate of 20kHz. The CoNNear model consists of three distinct modules: the cochlear stage ($\text{CoNNear}_{\text{cochlea}}$), IHC stage ($\text{CoNNear}_{\text{IHC}}$), and ANF stage. The ANF stage is subdivided into three different types of ANFs: $\text{CoNNear}_{\text{ANfH}}$, $\text{CoNNear}_{\text{ANfM}}$ and $\text{CoNNear}_{\text{ANfL}}$ for the HSR, MSR and LSR ANFs, respectively. The responses of the three ANF types are combined together to yield the final summed AN response r_F, by using weights H_{NH}, M_{NH} and L_{NH} that correspond to the number of HSR, MSR and LSR fibers in a NH periphery (H_{NH} = 13, M_{NH} = 3 and L_{NH} = 3 as reported in Verhulst et al. [17]). The $\text{CoNNear}_{\text{cochlea}}$, $\text{CoNNear}_{\text{IHC}}$ and $\text{CoNNear}_{\text{ANf}}$ modules comprise encoder-decoder CNN architectures that can be backpropagated through, thus facilitating the development of individualized audio-enhancement methods.

2.2 DNN-Based CS-Compensating HA Algorithms

From the NH CoNNear model, we can obtain a HI CoNNear model by retraining the $\text{CoNNear}_{\text{cochlea}}$ stage via transfer learning to simulate OHC loss [19], and by changing the weights of the different types of ANFs in the $\text{CoNNear}_{\text{ANf}}$ stage to model AN fiber loss, related to CS. The CoNNear HI periphery model can be individualized based on frequency-dependent degrees of OHC loss and CS. The individualized degree of CS and OHC loss of a listener can be obtained from diagnostic measurements using the rectangular amplitude-modulated envelope-following responses (RAM-EFRs) and distortion-product otoacoustic emissions (DPOAEs), respectively [20-21].

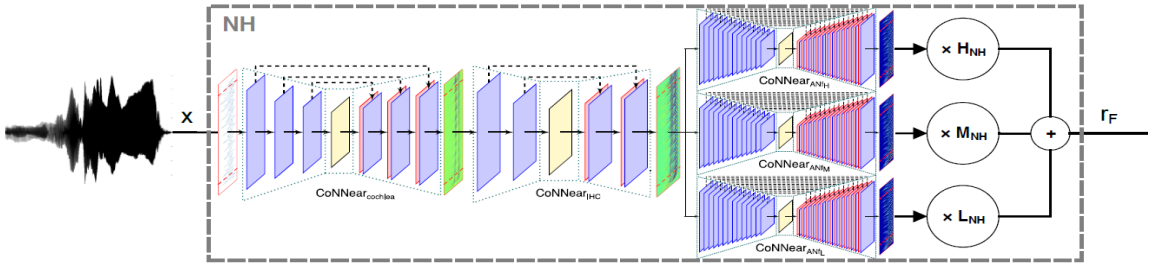

Figure 1. CoNNear model of the NH auditory periphery [12].

Here, based on the reference NH model and a HI CoNNear model, we use backpropagation to design ML-based audio signal processing algorithms that optimally compensate for CS [12-13]. As training dataset, we used 2310 randomly selected recordings from the TIMIT speech corpus [22]. The training procedure for a deep-neural-network- (DNN) based HA algorithm has been explained by Drakopoulos et al. [12-13] and goes as follows: A DNN-HA model is trained to process the input speech x into $\hat{x}$ such that the difference between the NH CoNNear model response r_F and the HI CoNNear model response $\hat{r}_F$ is minimized. Different CS-compensating HA algorithms can be designed in this way by defining a different set of loss functions [12-13]. These functions can focus on minimizing different aspects of the AN responses (e.g. free training, using more or less cochlear channels, limiting the frequency range, time-domain and frequency representations, also summed across CFs). Thanks to the modular nature of CoNNear, the loss functions can be fine-tuned for each of its distinct modules to optimally compensate for hearing impairment in each auditory stage.

In this work, we trained three DNN-HA models (CS_PReLU, CS_tanh_freq and CS_tanh) to compensate for a hearing-impairment with a CS profile of $H_{HI} = 7$, $M_{HI} = 0$ and $L_{HI} = 0$ ANFs, and no OHC loss. A CNN encoder-decoder architecture was used that comprised 16 layers (8 in the encoder and 8 in the decoder) as described in [12]. The three HA models were trained using different activation functions between the convolutional layers, and different sets of loss functions that were defined to minimize a combination of different aspects of the AN responses. The individual weights of the components in each joint loss function are listed in Tab. 1. A more detailed explanation of the different components of the loss functions that can be used during training is given by Drakopoulos et al. [13].

Table 1. Loss function weights and activation function used per DNN-HA model.

Loss function weights	**CS_PReLU**	**CS_tanh_freq**	**CS_tanh**
Low-frequency CFs emphasized	No	Yes	No
AN population response	High	Low	High
Complex STFT AN response per channel	High	Low	High
Squared complex STFT	Yes	No	Yes
Offset removal of resting firing rate AN	Yes	No	Yes
Non-linear activation function	PReLU	tanh	tanh

The CS_PReLU and CS_tanh HA models were trained using the same weights across their different loss components, they only differ in the used non-linear activation function. In all three HA models, the time-domain AN responses were squared (loss function l_r² in [13]) and only AN responses above a certain threshold were minimized (threshold T_r in [13]). The squaring of the time-domain responses was done to emphasize the temporal contrast of the speech envelope modulation to focus the optimization on the enhancement of the most excited regions, since temporal envelope coding is essential for robust speech intelligibility [23-24]. The threshold T_r was applied in the loss functions to further focus on the temporal peaks of the responses, and hence avoid minimizing differences in the resting firing rates of the AN responses. In the CS_PReLU and CS_tanh models, a high weighting was used for the loss function of the AN population response and for the loss function of the complex STFT AN response, while this weight was set lower in CS_tanh_freq. The CS_PReLU and CS_tanh models both included a loss function for the squared complex STFT, and a loss function to remove the offset of the resting firing rate of the AN, while this was not the case for the CS_tanh_freq model. Only in the CS_tanh_freq model, a frequency weighting was applied to emphasize the processing of the low-frequency CFs (freq.emphasis in [13]), so that the high frequencies were processed less than the low. This could result in a more accurate optimization, since the corpus contains energy mostly at low frequencies and could lack information on how the processing needs to be done at high frequencies. Emphasizing the low CFs in the optimization might hence achieve better benefits in speech intelligibility.

2.3 HA Model Evaluation

The three trained HA models were evaluated on their ability to compensate for the considered CS profile of $H_{HI} = 7$, $M_{HI} = 0$ and $L_{HI} = 0$ ANFs. Post-mortem data from recent temporal-bone studies have shown that NH people have lost more than half of their AN innervations after the age of 50, therefore we chose this CS profile of severe AN fiber loss [25-26].

Speech stimuli were processed with the three trained DNN-HA models, to evaluate the auditory feature restoration capabilities of the ML algorithms using auditory model simulations. The DNN-HA processed stimuli were given to the HI CoNNear model with the considered CS profile, in order to compare the simulated AN responses of the NH CoNNear model to the responses of the HI CoNNear model, with and without applying the HA processing. This way, we could investigate the difference in responses between the NH and HI models, and see how the HA processing affects the

output for the HI case, aiming to restore the AN responses to the NH level. In this work, we present the processing outcomes for the words 'David draagt' (English: 'David carries'), extracted from the Flemish Matrix corpus, presented at a level of 70 dB SPL relative to the reference pressure $p_0 = 2 \cdot 10^{-5}$Pa, with a sampling frequency of 20kHz [27]. The DNN-HA models require an input that is a multiple of 256 samples, hence zero-padding was applied at the end of the 'David draagt' stimulus. For the input stimulus 'David draagt', we analyzed several metrics obtained from the simulated CoNNear AN responses. The first of the four presented metrics is the excitation pattern at the level of the basilar membrane, reflecting the root-mean-square (RMS) over time of the vibration of the BM per CF. The BM excitation patterns shows the vibration amplitude of the BM in function of the CFs along its length, in response to the full input stimulus 'David draagt'. The second presented metric is the excitation pattern at the level of the AN, reflecting the RMS over time of the summed AN response, which is the summation of the simulated firing rates of the HSR, MSR and LSR ANFs, each weighted by their respective number of fibers present (for the HI CoNNear model: $H_{HI} = 7$, $M_{HI} = 0$ and $L_{HI} = 0$; for the NH CoNNear model: $H_{NH} = 13$, $M_{NH} = 3$ and $L_{NH} = 3$), per CF. The third presented metric is the wave-1 response, which is the summation of the summed AN response across the different CFs in time, calibrated in order to match experimentally recorded wave-1 amplitudes. And the last presented metric is the AN summed response difference before and after HA processing, showing the intensity of AN firing rate per CF in time.

3. RESULTS

We evaluated the auditory feature restoration capabilities of the three trained ML-based HA algorithms, using auditory model simulations as described in the Methods. At the conference, we will present the processing outcomes for the speech stimulus to investigate which auditory features the different HA processing algorithms focused on to compensate for CS.

4. ACKNOWLEDGMENTS

Work supported by ERC-StG 678120 and EIC-Transition-101058278.

5. REFERENCES

[1] S.G. Kujawa and M.C. Liberman, "Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss," *Journal of Neuroscience,* 29(45):14077–85, 2009.

[2] A.C. Furman, S.G. Kujawa and M.C. Liberman, "Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates," *Journal of Neurophysiology*, 110(3):577–86, 2013.

[3] E. Lobarinas, C. Spankovich and C.G. Le Prell, "Evidence of "hidden hearing loss" following noise exposures that produce robust TTS and ABR wave-I amplitude reductions," *Hearing Research,* 349:155–63, 2017.

[4] H.M. Bharadwaj, S. Verhulst, L. Shaheen, M.C. Liberman and B.G. Shinn-Cunningham, "Cochlear neuropathy and the coding of supra-threshold sound," *Frontiers in Systems Neuroscience*, 8(26), 2014.

[5] A.M. Mepani, S. Verhulst, K.E. Hancock, M. Garrett, V. Vasilkov, K. Bennett et al., "Envelope following responses predict speech-in-noise performance in normal-hearing listeners," *Journal of Neurophysiology*, 125(4):1213–22, 2021.

[6] C.J. Plack, D. Barker and G. Prendergast, "Perceptual consequences of "hidden" hearing loss," *Trends in Hearing,* 18:1–11, 2014.

[7] Y. Sergeyenko, K. Lall, M.C. Liberman and S.G. Kujawa, "Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline," *Journal of Neuroscience*, 33(34):13686–94, 2013.

[8] S. Keshishzadeh, M. Garrett, V. Vasilkov and S. Verhulst, "The derived-band envelope following response and its sensitivity to sensorineural hearing deficits," *Hearing Research*, 392:107979, 2020.

[9] R. Drullman, J. Festen and R. Plomp, "Effect of reducing slow temporal modulations on speech reception," *Journal of the Acoustical Society of America*, 95:2670–80, 1994.

[10] G. Rance, C. McKay and D. Grayden, "Perceptual Characterization of Children with Auditory Neuropathy," *Ear and Hearing*, 25(1):34–46, 2004.

[11] F. Drakopoulos, V. Vasilkov, A. Osses Vecchi, T. Wartenberg and S. Verhulst, "Model-based hearing-restoration strategies for cochlear synaptopathy pathologies," *Hearing Research*, 424:108569, 2022.

[12] F. Drakopoulos and S. Verhulst, "A differentiable optimisation framework for the design of individualized DNN-based hearing-aid strategies," in *ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* p. 351–5, 2022.

[13] F. Drakopoulos and S. Verhulst, "A Neural-Network Framework for the Design of Individualised Hearing-Loss Compensation," in *IEEE/ACM Transactions on*

Audio, Speech, and Language Processing, vol. 31, pp. 2395-2409, 2023.

[14] D. Baby, A. Van Den Broucke and S. Verhulst, "A convolutional neural-network model of human cochlear mechanics and filter tuning for real-time applications," *Nature Machine Intelligence*, 3(2):134–43, 2021.

[15] F. Drakopoulos, D. Baby and S. Verhulst, "A convolutional neural-network framework for modelling auditory sensory cells and synapses," *Communications Biology*, 4(1), 2021.

[16] S. Verhulst, D. Baby, F. Drakopoulos and A. Van Den Broucke, "A neural network model for cochlear mechanics and processing," WO2020249532, 2020.

[17] S. Verhulst S, A. Altoè and V. Vasilkov, "Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss," *Hearing Research*, 360:55–75, 2018.

[18] D.D. Greenwood, "A cochlear frequency‐position function for several species−29 years later," *Journal of the Acoustical Society of America*, 87:2592-2605, 1990.

[19] A. Van Den Broucke, D. Baby and S. Verhulst, "Hearing-impaired bio-inspired cochlear models for real-time auditory applications," in *Proc. of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, p. 2842–6, 2020.

[20] S. Keshishzadeh and S. Verhulst, "Individualized Cochlear Models Based on Distortion Product Otoacoustic Emissions," in *Proc. of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,* p. 403–7, 2021.

[21] S. Keshishzadeh, M. Garrett and S. Verhulst, "Towards Personalized Auditory Models: Predicting Individual Sensorineural Hearing-Loss Profiles From Recorded Human Auditory Physiology," *Trends in Hearing*, 25:1–22, 2021.

[22] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," *NASA STI/Recon technical report n*, vol. 93, p. 27403, 1993.

[23] A. Parthasarathy, E.L. Bartlett and S.G. Kujawa, "Age-related Changes in Neural Coding of Envelope Cues: Peripheral Declines and Central Compensation," *Neuroscience,* 407:21–31, 2019.

[24] V. Vasilkov, M. Garrett, M. Mauermann and S. Verhulst, "Enhancing the sensitivity of the envelope-following response for cochlear synaptopathy screening in humans: The role of stimulus envelope," *Hearing Research*, 400:108132, 2021.

[25] L. Viana, J. O'Malley, B. Burgess, D. Jones, C. Oliveira, F. Santos et al., "Cochlear neuropathy in human presbycusis: confocal analysis of hidden hearing loss in post-mortem tissue," *Hearing Research*, 327:78–88, 2015.

[26] P.Z. Wu, L.D. Liberman, K. Bennett, V. de Gruttola, J.T. O'Malley and M.C. Liberman, "Primary Neural Degeneration in the Human Cochlea: Evidence for Hidden Hearing Loss in the Aging Ear," *Neuroscience*, 407:8–20, 2019.

[27] R. Houben, J. Koopman, H. Luts, K.C. Wagener, A. van Wieringen, H. Verschuure and W.A. Dreschler, "Development of a Dutch matrix sentence test to assess speech intelligibility in noise," *International Journal of Audiology*, 53(10):760–3, 2014.