APPLYING BIOPHYSICAL AUDITORY PERIPHERY MODELS FOR REAL-TIME APPLICATIONS AND STUDIES OF HEARING IMPAIRMENT

Arthur Van Den Broucke[1] **Fotios Drakopoulos**[1] **Deepak Baby**[2] **Sarah Verhulst**[1]

[1] Dept. of Information Technology, Ghent University , Belgium
[2] Idiap Research Institute, Martigny, Switzerland
Arthur.VanDenBroucke@UGent.be, s.verhulst@UGent.be

ABSTRACT

Biophysically realistic models of cochlear processing are based on cascaded transmission-line (TL) models which capture longitudinal coupling, cochlear nonlinearities, as well as the human frequency selectivity. However, these models are slow to compute (in the order of seconds/minutes), explaining why less-accurate model descriptions of cochlear processing (e.g., gammatone, DRNL, MFCC) are still the standard for feature extractors or auditory front-ends. To overcome this gap, we present a hybrid approach in which convolutional neural network (CNN) techniques are combined with computational modelling to yield a real-time model of the human auditory periphery. A CNN was trained on speech corpus material to mimic a state-of-the-art biophysical model that can accurately represent the human cochlea and the ascending auditory pathway. The performance was compared against human data and simulations of the original model using basic stimuli (pure tones, clicks, etc.). Because the original peripheral model can simulate different degrees of sensorineural hearing loss, the normal-hearing CNN model can be adjusted in the same fashion to simulate hearing impairment. The neural-network character of these architectures allows for real-time, parallel and differentiable computations, which can serve in the next generation of hearing-aid and machine-hearing applications.

1. INTRODUCTION

Comparing state-of-the-art models of the human cochlea, a clear distinction is apparent between real-time models (e.g., gammatone, DRNL, MFCC) which are perceptually relevant but lack biophysically-relevant aspects, and nonlinear models (e.g., cascaded transmission-line (TL) models) which can achieve a high level of biophysical plausibility but are usually computationally expensive. To address this ever-present trade-off between physiological accuracy and computational speed, we designed a real-time cochlear model based on deep neural networks (DNNs). The resulting DNN-based model was able to grasp the relevant mechanics of the normal-hearing (NH) human cochlea in great detail and was also trained to include different types of hearing impairment. The proposed framework was applied to the successive stages of the auditory periphery to yield a fast-operating, biophysically-relevant, DNN-based model that can include different frequency-specific variations of outer-hair-cell loss and cochlear synaptopathy. The resulting models can have a great application value for real-time applications (machine hearing, next generation of hearings-aids) and in various studies of hearing-impairment. In this extended abstract, we summarize a couple of its features which are described in [1, 2].

2. METHODS

2.1 Normal Hearing CoNNear

2.1.1 Cochlea

The neural-network architectures that we adopted were encoder-decoder convolutional-neural-networks (CNNs), such as the one shown in Figure 1. We opted for an end-to-end, CNN-based model that can yield a real-time solution of the reference cochlear TL model [1]. The training, for the replacement model of the cochlear stage, was done on a GPU using 2310 combinations of speech inputs (TIMIT corpus) and their reference-model cochlear output [2], and lasted for approximately 2 days. Training consisted of updating the neural network parameters by minimizing the L1-loss between the predicted and target basilar membrane vibrations for the different frequency channels spanning the human frequency range (100 Hz - 12 kHz).

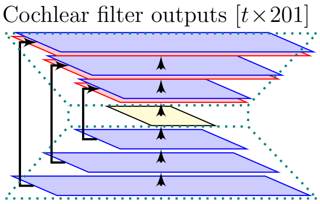

Figure 1. CoNNear-architecture for cochlear stage [2]: The CoNNear architecture is an auto-encoder, convolutional neural network framework which is connected using strided convolutions between layers, and skip-connections. It maps the time-domain sound input (bottom) to 201 time-domain cochlear filter outputs of different center frequencies (top). The depicted model has four encoding and decoding layers and uses a tanh activation function between the layers.

2.1.2 Inner-hair-cells and Auditory-nerve -fibers

The framework that we presented for the training of DNN-based cochlear models can also be applied to the next stages of the hearing pathway, i.e. the inner-hair-cell (IHC) and auditory-nerve-fiber (ANF) complex, in a similar fashion. Different hyperparameters were used depending on the nature of each stage to determine the optimal architectures. The IHC and ANF models can simulate single-unit responses or be connected to our cochlear model to simulate population responses.

2.2 Hearing Impairment

2.2.1 Outer-hair-cell loss

The reference auditory periphery model can be adjusted to include expressions of outer-hair-cell (OHC) deficits in its computing stages [3]. Hence the question was asked if we could obtain a hearing-impaired (HI) CoNNear model as well. To achieve this, we saw that the trained NH CoNNear model could serve as a starting point. By applying transfer learning [4], a technique where you start from an already trained network -in this case the NH CoNNear-, we were able to obtain a hearing-impaired cochlear model with only 50 additional, hearing-impaired, training utterances and 10 minutes of training on a GPU [4].

2.2.2 Cochlear Synaptopathy

The number of ANF used in the reference model can also be adjusted to simulate the innervations of a normal-hearing IHC or an IHC with auditory-nerve deficits (i.e. cochlear synaptopathy).

3. RESULTS

3.1 Cochlear Mechanics

To evaluate the performance of the trained CoNNear models, basic auditory stimuli, unseen during the training phase (i.e., clicks and pure tones), were given as inputs to address the performance of CoNNear on the level dependent tuning, the excitation patterns of basic auditory stimuli and the distortion product otoacoustic emissions of the cochlea. Based on the aforementioned properties of cochlear mechanics, we concluded [2, 4] that our trained models were able to grasp the cochlear properties in the same biophysically correct manner as the reference TL model.

3.2 Timing

The second aspect we investigated relates to the real-time character of the CoNNear models. As seen in Table 1, the reference TL model needs 25 seconds to process an audio fragment of 102.4 ms on a CPU. When the same fragment was given to CoNNear, calculations on a CPU were around 100 times faster. When using a GPU, CoNNear was 3000 times faster than the reference and we obtained latencies below 8 ms, which can be considered real-time.

	CPU	**GPU**
Transmission Line model	25.156 s	NA
NH CoNNear	0.236 s	0.0073 s
HI CoNNear	0.242 s	0.0079 s

Table 1. Computational time of trained CoNNear models: Execution time of the reference TL model compared to a trained version of the NH and HI CoNNear model for an audio fragment of 102.4 ms. Computations were performed on a CPU (Apple MacBook Air, 1.8 GHz Dual-Core processor) as well as on a GPU (NVIDIA GTX1080).

4. DISCUSSION

In this work, we departed from a slow-to-compute computational model of the auditory periphery to yield a real-time, CNN-based approximation. This was done for models of the first stages of human hearing (i.e., cochlea, inner-hair-cells, auditory-nerve-fibers) and this for both structures that resemble normal hearing profiles as well as models that include hearing-impaired expressions (outer-hair-cell loss, cochlear synaptopathy). In the future this framework can be extended to include more stages of the ascending auditory pathway such as the brainstem processing. The resulting models can provide a significant speed-up in computational time making this a viable option for real-time auditory applications, and can be included within backpropagation networks due to their differentiable nature. The trained HI CoNNear models can also be used in closed-loop applications to design hearing-loss compensation algorithms for the next-generation of hearing aids.

5. ACKNOWLEDGMENTS

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

6. REFERENCES

[1] S. Verhulst, A. Altoè, and V. Vasilkov, "Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss," *Hearing Res.*, no. 360, pp. 55–75, 2018.

[2] D. Baby, A. V. D. Broucke, and S. Verhulst, "A convolutional neural-network model of human cochlear mechanics andfilter tuning for real-time applications," *arXiv preprint arXiv:2004.14832*, 2020.

[3] S. Verhulst, A. Jagadeesh, M. Mauermann, and F. Ernst, "Individual differences in auditory brainstem response wave characteristics: Relations to different aspects of peripheral hearing loss," *Trends in hearing*, vol. 20, 2016.

[4] A. V. D. Broucke, D. Baby, and S. Verhulst, "Hearing-impaired bio-inspired cochlear models for real-time auditory applications," in *Proc. Interspeech 2020*, pp. 2842–2846, 2020.