DOI: 10.61782/fa.2023.1002

EVALUATION OF THE REVISED COMPUTATIONAL MODEL OF AUDITORY SIGNAL PROCESSING AND PERCEPTION

Lily Cassandra Paulick[1]* **Helia Relaño-Iborra**[1] **Torsten Dau**[1]

[1] Hearing Systems Section, Department of Health Technology,
Technical University of Denmark,
Kgs. Lyngby, Denmark

ABSTRACT

A revised version of the computational auditory signal processing and perception model (CASP, [1]) is proposed. CASP is a functional model designed for the prediction of data from various psychoacoustic conditions, including intensity discrimination, spectral and temporal masking and amplitude modulation detection and masking. The revised version of CASP proposed here introduces two main changes to the original model: (i) a more realistic, non-linear inner hair cell stage replacing the simplified linear processing stage of the original model and (ii) an updated back-end processing that captures off-frequency listening through an automatic channel selection algorithm. The effects of these two modifications in the model are analysed in various signal detection and masking conditions, including intensity discrimination, forward masking as well as modulation detection using narrowband noise carriers. Model predictions are compared across model configurations and examined in terms of the internal representations of the signals at different stages of auditory processing within the model. The potential implications of the revised signal processing in the model on predictions of effects of sensorineural hearing loss are discussed.

Keywords: *auditory modelling, psychoacoustics, inner hair cell transduction*

1. INTRODUCTION

The computational model of auditory signal processing and perception (CASP, [1, 2]) has been shown to be a successful predictor of listeners' behaviour in a wide range of psychoacoustic measurements, from intensity discrimination, spectral and temporal masking to modulation detection. It accounts for data from listeners with both normal hearing (NH) and hearing impairment (HI), explaining various aspects of simultaneous and non-simultaneous masking [1, 2]. CASP provides a powerful tool to test hypotheses and assess how various stages of the auditory pathway contribute to the overall processing and perception of the stimulus. Most auditory models consist of different stages that reflect the basic parts of the auditory system, and model the underlying processes with varying degrees of detail [3]. While these stages are theoretically independent, each stage determines the operating point of the succeeding ones and therefore, it is essential to understand how the different stages interact with each other and how deficits in certain parts of the system can influence the processing further along the auditory pathway. Sensorineural hearing loss is typically caused by damage to inner hair cells (IHC) and outer hair cells (OHC) as well as synaptic loss in the auditory nerve. While the effects of OHC loss are often included in cochlear models, the IHC stage requires separate processing stages.

In this study, the inner hair cell transduction was of particular focus. This stage was in previous versions of CASP implemented as a simple envelope extraction [1] consisting of a half-wave rectification and low-pass filtering. However, based on physiological recordings, it is known that the IHC transduction reflects a non-linear process, that includes a compressive behaviour towards

---

**Corresponding author*: *lpau@dtu.dk.*

higher sound pressure levels [4]. This non-linearity has been argued to be an essential part of cochlear processing [5]. A more realistic implementation of this stage in the overall CASP model is thus hypothesised to improve model predictions, in particular with regard to predictions of data from listeners with a hearing impairment. The large variability observed in data across listeners with hearing impairment, especially in speech-related tasks, is still largely underestimated by current modelling studies [6]. The present study aimed to validate the revised CASP model implementation in psychoacoustic conditions for listeners with normal hearing. An outlook is provided towards predicting behaviour of listeners with hearing impairment in fundamental pyschoacoustic conditions known to relate to measures of speech intelligibility, such as spectro-temporal masking [7].

2. MODEL DESCRIPTION

The model can be split into its preprocessing stages (that build the corresponding internal representation of the input stimulus) and the back-end processing that includes the decision device. The preprocessing stages of the model are shown in Fig. 1. The input signal is processed through an outer and middle ear filter, a dual-resonance non-linear filterbank (DRNL, [8]), followed by the IHC model, adaptation loops and, finally, a modulation filterbank. The output of these preprocessing stages is a three-dimensional internal representation with dimensions time, auditory frequency and modulation frequency. The back-end processing of the model consists of an optimal detector [9] that is designed to be used in connection with n-alternative forced choice (nAFC) paradigms that adaptively track the thresholds in different psychoacoustic tasks. Here, the resolution is limited by internal noise added to the input to the optimal detector. The revised model implementation introduces modifications primarily at the inner hair cell stage and in the back-end processing, as described in greater detail below.

2.1 Inner hair cell model

A major element of the revised model implementation is the inclusion of a more realistic IHC transduction model. Previously, this stage was functionally implemented as a simple half-wave rectification and low-pass filtering stage [1, 9]. While this simplified model is able to account for the loss of phase-locking towards higher frequencies, it does not account for the compressive input-output function known from mammalian physiological recordings [4]. Instead, a physiologically inspired model is proposed here as a revised component of the overall CASP model. The IHC model implementation closely follows that of previous biophysical studies [10, 11]. It consists of two stages: first, a fitted Boltzmann function converts stereocilia displacement to mechano-electrical conductance. Then, the conductance is used to calculate the receptor potential through an electrical circuit analogy. The receptor potential of the inner hair cell represents the output of the IHC model. Both stages follow the implementation and parameter fitting of [11].

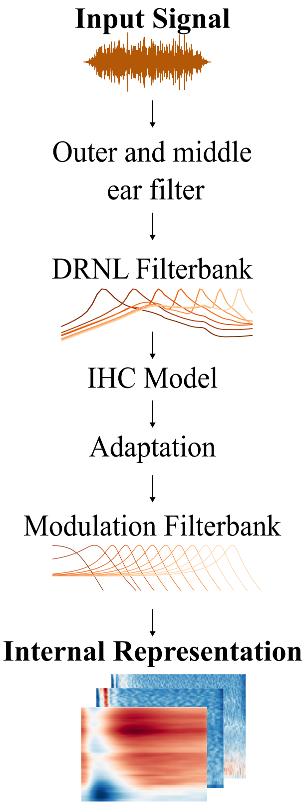

Figure 1. Preprocessing stages of the CASP model. The input consists of a time signal, while the output consists of a three-dimensional internal representation.

2.2 Frequency channel selection

An additional element of the revised CASP is the automatised process of selecting the relevant peripheral channels to be included in the detection process. In the original

CASP, the region of interest needed to be selected manually before each experimental run based on prior knowledge and assumptions. Here, an automatic channel selection is proposed that allows the model to decide on the relevant region(s) of analysis without any *a priori* knowledge. The selection is based on the internal representation in the auditory frequency domain (i.e., before the modulation filterbank stage). The power is calculated in each frequency band, after which the frequency channel with the highest power is determined and the region of analysis is defined around that channel by including all neighbouring channels for which less than a 25% power drop from the maximum to each side is detected.

3. RESULTS

Simulations of a forward masking paradigm were carried out using the same paradigm as in [1]. Threshold curves were found for the detection of a 4 kHz tone played after broadband noise at three different levels (40, 60 and 80 dB). In Fig. 2, the simulation results are shown using a predetermined frequency region from 2 to 8 kHz for the model, i.e. no channel selection was applied. Here, the green function represents the simulations using the linear IHC model, and the blue function represents the simulations using the non-linear IHC model. Human data [1] are shown in black alongside the simulations. Both model predictions account well for the data, with a slightly steeper decline observed in the threshold function predicted with the non-linear IHC model.

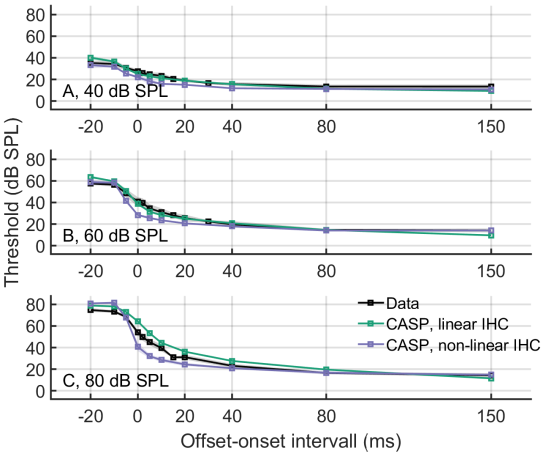

Figure 2. Model predictions using a predetermined frequency range from 2 to 8 kHz for CASP with the linear IHC model (green) and the non-linear IHC model (blue) are shown alongside the data (black, [1]) for the three different forward masking experiments.

Similarly, Fig. 3 shows simulation results alongside the human data from [1] using the proposed channel selection algorithm. The green function represents again the simulations using the linear IHC model, and the blue function represents the simulations obtained with the non-linear IHC model. The results show that, with the proposed frequency channel selection, the two model configurations with the linear and the non-linear IHC model are still able to account for the data without making use of prior knowledge of the target frequency to determine the region of interest. Tab. 1 shows error metrics for the four different model configurations, corresponding to combinations of the IHC model and the channel selection algorithm. The mean absolute error (MAE) and the correlation coefficient (ρ) between the model predictions and the data are provided in each case. The values are averaged across the three different conditions, corresponding to the different noise levels. All four models show comparable results and a good agreement with the data. While the

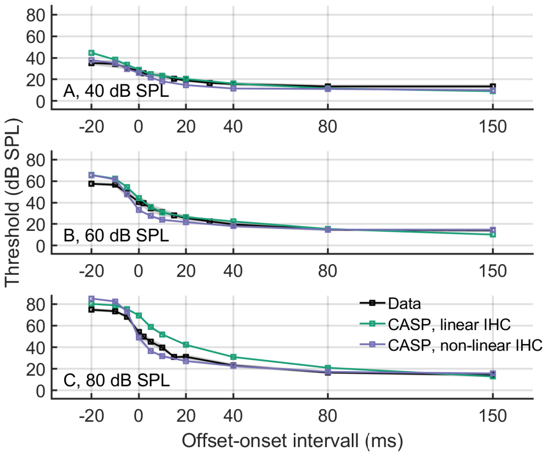

Figure 3. Model predictions using automatic channel selection for CASP with the linear IHC model (green) and the non-linear IHC model (blue) are shown alongside the data (black, [1]) for the three different forward masking experiments.

model predictions obtained with the linear IHC model and the predetermined channel region showed the best agreement with the data, with an MAE = 3.3 dB and a correlation of $\rho = 0.99$, the predictions with the non-linear IHC model still showed results on par with the previous version of the model, with an MAE = 5 dB and a correlation of $\rho = 0.96$. Moreover, a marginal improvement can be seen in the model performance obtained with the non-linear IHC model when applying the channel selection, as the non-linear model enhances across-frequency cues that are not necessarily reflected in the predetermined channel region.

Table 1. Error metrics for the predictions with different model configurations.The first value represents the mean absolute error (MAE), and the second value the correlation coefficient (ρ) between the predictions and data. In each case, the values have been averaged across the three conditions.

	linear IHC	non-linear IHC
predetermined channels	MAE = 3.3 dB; $\rho = 0.99$	MAE = 5 dB; $\rho = 0.96$
channel selection	MAE = 4.8 dB; $\rho = 0.99$	MAE = 4.1 dB; $\rho = 0.98$

4. CONCLUSION

A revised model implementation for the computational model of auditory signal processing and perception was presented. The model was validated through predictions of data from listeners with normal hearing in forward masking tasks (shown here), as well as other psychoacoustic conditions (not explicitly shown here), providing performance levels equivalent to those obtained with previous versions of the model. The preprocessing stages for the revised CASP model may prove valuable for the prediction of data from listeners with hearing impairment, where the inclusion of a more realistic inner hair cell transduction is hypothesised to improve individual predictions, especially in speech-related tasks.

5. REFERENCES

[1] M. L. Jepsen, S. D. Ewert, and T. Dau, “A computational model of human auditory signal processing and perception,” *The Journal of the Acoustical Society of America*, vol. 124, pp. 422–438, 7 2008.

[2] M. L. Jepsen and T. Dau, “Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss,” *The Journal of the Acoustical Society of America*, vol. 129, pp. 262–281, 1 2011.

[3] A. O. Vecchi, L. Varnet, L. H. Carney, T. Dau, I. C. Bruce, S. Verhulst, and P. Majdak, “A comparative study of eight human auditory models of monaural processing,” *Acta Acustica*, vol. 6, p. 17, 5 2022.

[4] P. Dallos, “Response characteristics of mammalian cochlear hair cells,” *The Journal of Neuroscience*, vol. 5, pp. 1591–1608, 1985.

[5] L. H. Carney, “Supra-threshold hearing and fluctuation profiles: Implications for sensorineural and hidden hearing loss,” 8 2018.

[6] H. Relaño-Iborra and T. Dau, “Speech intelligibility prediction based on modulation frequency-selective processing speech intelligibility prediction based on modulation frequency-selective processing,” *Hearing Research*, vol. 426, p. 108610, 2022.

[7] J. Zaar, L. B. Simonsen, T. Dau, and S. Laugesen, “Toward a clinically viable spectro-temporal modulation test for predicting supra-threshold speech reception in hearing-impaired listeners,” *Hearing Research*, vol. 427, 1 2023.

[8] E. A. Lopez-Poveda and R. Meddis, “A human nonlinear cochlear filterbank,” *The Journal of the Acoustical Society of America*, vol. 110, pp. 3107–3118, 12 2001.

[9] T. Dau, D. Püschel, and A. Kohlrausch, “A quantitative model of the “effective” signal processing in the auditory system. i. model structure,” *The Journal of the Acoustical Society of America*, vol. 99, pp. 3615–3622, 6 1996.

[10] E. A. Lopez-Poveda and A. Eustaquio-Martín, “A biophysical model of the inner hair cell: The contribution of potassium currents to peripheral auditory compression,” *JARO - Journal of the Association for Research in Otolaryngology*, vol. 7, pp. 218–235, 9 2006.

[11] A. Altoè, V. Pulkki, and S. Verhulst, “Model-based estimation of the frequency tuning of the inner-hair-cell stereocilia from neural tuning curves,” *The Journal of the Acoustical Society of America*, vol. 141, pp. 4438–4451, 6 2017.