Modeling and clinical diagnosis of dead regions in the cochlea
Warnaar, B.

Citation for published version (APA):
Warnaar, B. (2013). Modeling and clinical diagnosis of dead regions in the cochlea

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Appendix A

THE CASP MODEL
Figure A.1. The CASP model redrawn from Jepsen et al. (2008), with permission from the authors. The model accepts digital waveforms as input, and attempts to simulate a psychophysical response of a normal-hearing listener to a detection task.
A.1 The CASP model

The computational auditory signal-processing and perception (CASP) model that was developed by Jepsen et al. (2008) was used to simulate the model results described in Chapters 3, 4 and 5. The overall structure of the CASP model is redrawn in Fig. A.1. The model consists of a chain of modules, each simulating a functional operation of the auditory system. Several modules can be related to a specific part of the auditory system, while other modules provide a more functional description of the auditory processing. The aim of the chain of modules is to compute human auditory signal-processing and perception. The model is able to simulate the psychophysical responses to (a variety of) digitalized auditory inputs.

This appendix provides an overview of the CASP model as described by Jepsen et al. (2008). Firstly, details of the model are described per module with examples of signals used in this thesis. Secondly, instructions are given on how to obtain and use the model. With these instructions, the reader should be able to reproduce normal-hearing model data that were presented in this thesis, i.e., Figs. 3.3, 3.4, 4.1 (panel A), 4.4, and 5.1 (panel A).

Where applicable, additional model details are described that can be used to simulate the hearing-impaired data as presented in this thesis.

A.1 Module details

A.1.1 Outer and middle-ear transformations

The input to the CASP model is a digital signal with an amplitude between 0 and 1. An input of 0 corresponds to a sound level of 0 dB SPL. An input of 1 corresponds to the maximum sound level of 100 dB SPL for which the model is designed. The outer and middle-ear module scales the digital sound input to a physical pressure waveform in Pascal (Pa). The pressure wave is used as input to spectral transformations of the headphone-to-eardrum and the mechanical impedance of the middle-ear ossicles.

The transformations are interpolated with a linear-phase 512-order finite impulse response (FIR) filter to empirical data. Outer-ear data was taken from Pralong and Carlile (1996). This study measured headphone-to-eardrum transfer functions by inserting a custom-made microphone deep within the ear canal of human listeners. Middle-ear data was taken from Goode et al. (1994). This study measured the stapes peak-to-peak displacement in animal cadavers. The combined outer and middle-ear transformation function is shown in Fig. A.2.

A.1.2 Dual-resonance non-linear filterbank

The DRNL filterbank simulates the stapes motion in meters per second (m/s) to basilar membrane motion in meters per second (m/s) at various characteristic frequencies (CF). A channel of the DRNL consists of two paths, a linear path and a non-linear path. It is tempting to think of the linear path as the passive BM response (IHCs) and the non-linear path as the active mechanisms in the cochlea (OHCs). However, this is incorrect. Combined the paths simulate the
Figure A.2. In this thesis, PTC and TEN maskers were used in combination with pure-tone probes. The functional behavior of the outer and middle-ear were simulated by linear spectral transformations.

The linear path consists of a linear gain $g$, a cascade of two gammatone bandpass filters with center frequency $CF_{lin}$ and bandwidth $BW_{lin}$, and a cascade of four 2nd-order lowpass Butterworth filters with cutoff frequency $LP_{lin}$. The linear path dominates the I/O function at level above 70-80 dB. The non-linear path consists of a cascade of three 1st-order gammatone filters with characteristic frequency $CF_{nl}$ and bandwidth $BW_{nl}$, a non-linear gain that is compressive between 40 and 70 dB SPL, a second cascade of three 1st-order gammatone filters with similar parameters as the first set of gammatone filters, and finally a cascade of three 2nd-order lowpass Butterworth filters with cutoff frequency.
A.1 The CASP model

outer-, and middle-ear transforms

Figure A.3. The DRNL filterbank consists of a linear (blue) and non-linear (purple) path. The input of the filterbank is analyzed in the level and frequency domains at various frequencies. Panels (A) and (B) show the level and frequency response of a DRNL channel with a characteristic frequency of 1 kHz.

LP_{nl}. The compressive I/O function of the non-linear path, which is also referred to as a broken-stick non-linearity, is given by:

\[
y(t) = \text{sign}[x(t)] \cdot \min[a|x(t)|, b|x(t)|^c]
\]  
(A.1)
where \( x(t) \) and \( y(t) \) are the input and output signals of the non-linear gain in the non-linear path, and \( a \), \( b \) and \( c \) are model parameters. All parameters in the linear-, and non-linear path were fitted at the frequencies 250, 500, 1000, 2000, 4000 and 8000 Hz, and interpolated between these frequencies.

The DRNL filterbank was first described by Meddis et al. (2001). In this study, the filterbank was fitted to animal data. Lopez-Poveda and Meddis (2001) modified the filterbank to describe pulsation-threshold data estimating peripheral compression in the human cochlea (Plack & Oxenham, 2000). Finally, the parameters were modified by Jepsen et al. (2008) to improve the simulation of compression above 1.5 kHz. This last adjustment was motivated by studies of Lopez-Poveda et al. (2003) and Rosengard et al. (2005).

The hearing-impaired data that is presented in this thesis were simulated by the CASP model with modifications in the DRNL filterbank. Regions of hair cell impairment were characterized by a frequency range and an operation to the I/O function within this region. The DRNL channels that were affected had a CF that corresponded to the frequency range of impaired region. The cutoff frequency of channels was determined by edge frequency of the impaired region. The loss of IHCs was simulated by the removal of channels inside the affected frequency range. Regions with IHC insensitive were simulated by an attenuation of the channel output equal to the level of IHC attenuation. Dysfunction of OHCs was simulated by reducing the parameter \( a \) in the non-linear path by an amount equal to the loss. As an example, a high-frequency dead region (IHC loss) with edge frequency \( F_e \) equal to 3 kHz and an OHC dysfunction causing an elevation of the absolute threshold of -3 dB at frequencies above 2.5 kHz was simulated by removing DRNL channels with CF above 3 kHz and multiply the parameter \( a \) in the non-linear gain function of the non-linear path by 0.5 in channels between 2.5 and 3 kHz. The parameter values that were used to simulate individual listener data can be found in Tables 4.1 and 5.1 of this thesis.

Furthermore, the channel density and range of channels were adjusted for the simulations presented in this thesis. The density of channels was increased from 1/ERB to 8/ERB to improve off-frequency listening in the model. The range of channels was set to at least -1 to +1 octaves centered on the frequency of interest. In case of normal-hearing simulations, the frequency of interest corresponded to the frequency of the probe. In case of simulating dead regions, this frequency corresponded to the edge frequency of the dead region, \( F_e \). In case of simulation of IHC insensitivity, the range included the cutoff frequency of insensitive IHCs and the probe frequency with an extra octave of channels towards lower- and higher frequencies.

### A.1.3 Hair Cell Transduction

The hair cell transduction module simulates the transduction of basilar membrane motion into receptor potentials of the auditory nerve. The module consists of a half-wave rectification and a lowpass filter with a cutoff frequency of 1 kHz. Half-wave rectification removes the negative part of the input signal by
A.1.4 Expansion

The expansion module performs a squaring of the signal. It is motivated by rate-versus-level functions of the auditory nerve fibers near threshold (Yates et al., 1990; Muller et al., 1991).

A.1.5 Adaptation

Adaptation to input was found in the receptor potentials of the auditory nerve (Smith, 1977; Westermann & Smith, 1984). The adaptive properties of the auditory periphery, i.e., time-dependent level changes in the input are simulated.
in the adaptation module. The module consists of a sequence of five feedback loops, each with a capacitor with a different time constants varied between 5 and 500 ms. The I/O function of the sequence of loops to a steady state signal (a long duration compared to the time constants of the capacitors) is given by:

\[ O_{t \to \infty} = 2n \cdot \sqrt{I} = 10 \cdot \sqrt{I} \]  \hspace{1cm} (A.2)

where \( n \) is the number of loops. Changes in input are transformed to the output with a delay, because the capacitors need to be charging or discharging to reach a steady state transformation. These delayed effects allow the model to simulate forward masking effects as described by Hanna et al. (1982) and Zwicker (1984).

The version of the adaptation model that was used in the CASP model was a slightly modified version (Münkner, 1993) of the adaptation model as described by Püschel (1988). The modification reduced the onset response amplitude to 10 times the output of the same signal presented long enough to fully charge the capacitors (steady state).

A.1.6 Modulation filterbank

The modulation filterbank module is a linear filterbank that simulates the detection of modulation and masking of modulation in the signal. The modulation filterbank filters amplitude changes in the envelope of the signal. The filters in the filterbank are parameterized by a 5 Hz bandwidth below 10 Hz. The lowest filter has a peak response at 0 Hz and a cutoff frequency at 2.5 Hz. From 10 to 1000 Hz the modulation filters are centered at logarithmic intervals with a bandwidth equal to a quality factor (\( Q \), see Eq. (3.3)) of 2.

A.1.7 Internal noise and optimal detector

The psychophysical response of a listener is simulated by an internal noise and a detector that optimizes the detection of a target stimulus. The internal noise was calibrated such that it provided sufficient noise to discriminate a 1-dB level change in a long-duration signal with fixed frequency and a level of 60 dB SPL.

Detection is determined based on internal model representations of reference stimuli (masker only), the signal stimulus (masker plus probe) and a target stimulus, which is referred to as the template. Internal model representations in the optimal detector are a multi-dimensional function of time (waveform), frequency (DRNL channel) and modulation frequency (modulation filterbank channel) in model units (MU), as is shown for the different stimuli in Fig. A.5.

The template is constructed by averaging and normalizing ten different internal representations of masker plus a suprathreshold probe minus representations of maskers only. Using maskers to create the template is necessary, because internal representations change with level and may be affected by non-linear processing of the DRNL filterbank. The template is cross-correlated with the internal representation of the reference stimuli and the signal stimulus. If the difference between these correlations exceeds a predefined criterion, in favor of the signal stimulus, the signal is detected. The criterion can be expressed in
A.1 The CASP model

Figure A.5. The output of the modulation filterbank is a 3-dimensional model representation of the signal. The optimal detector creates a template of the target signal, and representation of test stimuli. The output of the model (detected or not detected) is calculated by cross-correlation of the template with the signal + reference - reference (probe + masker - masker). An internal noise is added to all stimuli to limit the resolution of the model.

terms of a probability of detection. Using the theory of signal detection (Green & Swets, 1966), the probability in the CASP model for a 3AFC procedure is given by:

\[
P(\text{correct}) = 1 - \text{erfc}
\left(\frac{0.765 \cdot \mu}{\sqrt{N}} - 0.423 \cdot 0.7071068\right)/2
\]  

(A.3)

where, \(P(\text{correct})\) is the probability of finding the probe, \(\mu\) is the difference be-
between the mean of the cross-correlated internal representations of the signal stimulus with the template and the largest cross-correlated internal representation of the reference stimuli with the template, \( N \) is the internal noise, \( \sigma^2 \) is the variance of the internal noise and \( \text{erfc} \) is the error function given by:

\[
\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt
\]

The simulations of dead regions presented in this thesis were performed with a limited frequency range of DRNL channels. This generally lowered the number of DRNL channels. The increased density of DRNL channels to 8/ERB also changed the number of DRNL channels used to simulate listener data. The normalization of the cross-correlated internal representations was adjusted accordingly to the number of channels (DRNL and modulation channels) to simulate with a constant internal noise. The normalization was adjusted according to:

\[
\mu_{\text{normalization}} \rightarrow \mu_{\text{normalization}} \cdot \sqrt{\frac{F}{F_{\text{base}}} \cdot \sqrt{\frac{D_{\text{base}}}{D}}}
\]

where, \( F \) and \( D \) are, respectively, the frequency range and density of DRNL channels, \( F_{\text{base}} \) and \( D_{\text{base}} \) are, respectively, the frequency range and density of DRNL channels that were used to calibrate the internal noise of the model.

Absolute thresholds were simulated in the CASP model by restricting the input of the adaptation loops to a minimum limit. Input that was presented at a level below the minimum was increased to the minimum level. This created an internal representation of stimuli, that were presented below the absolute threshold, that could not be discerned from stimuli presented at absolute threshold level. Since a reference stimulus was always subtracted from the signal stimulus, any information below threshold was not used to detect the signal.

\section*{A.2 How to use the CASP model}

\subsection*{A.2.1 Access to the source code}

The Auditory Modelling Toolbox is an initiative of several research groups in audiology to make software code for auditory tools available to a wider audience. Researchers interested in using the tools can access the toolbox through a web portal on the internet. Source code of the CASP model, and other tools, can be downloaded. Instruction and examples on how to use the model can also be downloaded from the same website.

Web portal: \textit{http://amtoolbox.sourceforge.net/}
CASP model: \textit{http://sourceforge.net/projects/caspmodel}

\subsection*{A.2.2 Using basic parameters}

As explained above, the CASP model contains many parameters. Most of these parameters were fitted to simulate normal-hearing auditory performance and
are included in the source code package of the CASP model. However, several parameters need to be considered when simulating listener data. The non-linear performance of the model and the extensive operations of the model influence how the model is used to simulate listener data correctly.

Firstly, the internal representation of stimuli changes non-linearly with level, because of the non-linear I/O function of the DRNL filterbank. This means that the suprathreshold level of the signal used to create a template changes the cross-correlation values between template and reference and signal stimuli. The simulated masking thresholds differ when, for example, the template is created with a probe presented at +30 or +60 dB SL. The optimal sensation level for the construction of a template is between 20 and 30 dB above the masking threshold. Typically, PTCs with steep slopes and simulations of dead regions require iterative changes of initial test conditions to simulate listener data accurately.

Secondly, simulation time increases linearly with the number of DRNL channels. The simulation time of each DRNL channel is noticeable, because DRNL channels are decomposed into several modulation filterbank channels. The creation of a template requires, preferably, ten runs, in which a target stimulus and a reference stimulus are simulated. Every interval of the 3AFC procedure needs another run of the model, and the complete model is run multiple times for each decision. This means that the simulation of a masking threshold for a single test condition may take five minutes, even when using a modern personal computer (at time of writing, 2013). The simulation of a set of PTCs with repetitions to improve accuracy may easily take several hours to a day of simulation time. Therefore, it is recommended to estimate the frequency range of interest before starting simulations, which may require some ‘Fingerspitzengefühl’.