
Stephan D. Ewert*, Steffen Kortlang and Volker Hohmann

*Corresponding author’s address: Medizinische Physik, Universität Oldenburg, Carl-von-Ossietzky Str. 9-11, Oldenburg, 26111, Niedersachsen, Germany, stephan.ewert@uni-oldenburg.de

In hearing aids, amplification and dynamic range compression typically aim at compensating the deficits associated with outer-hair cell (OHC) loss. Nevertheless, success shows large inter-individual variability and hearing-impaired listeners generally still have considerable problems in complex acoustic communication situations including noise and reverberation. These problems could be related to inner-hair cell (IHC) damage and reduced frequency selectivity resulting in a loss of spectro-temporal coding fidelity. Here a model-based, fast-acting dynamic compression algorithm which aims at approximating the normal-hearing BM input-output function in hearing-impaired listeners is suggested. The algorithm is fitted by estimating low-level gain loss (OHC loss) from adaptive categorical loudness scaling data and audiometric thresholds based on Ewert and Grimm [in: proc. ISAAR (2012)] and Jürgens et al. [Hear. Res. 270, 177 (2011)]. Aided speech intelligibility was measured in stationary and fluctuating noise and related to the estimated OHC loss. To improve diagnostics of OHC and IHC loss, a series of five psychoacoustic measurements was conducted aiming at a direct quantification of IHC damage in a group of six young and elderly normal-hearing and twelve hearing-impaired listeners. A model is suggested to account for the temporal fine-structure detection and discrimination data. Work funded by BMBF 01EZ0741 and DFG FOR1732.

Published by the Acoustical Society of America through the American Institute of Physics
MODEL-BASED DYNAMIC COMPRESSION

The proposed physiologically motivated, model-based dynamic compression algorithm (MDC2) is based on an earlier version suggested in Ewert and Grimm (2011). It aims at restoring the normal-hearing basilar membrane input-output function in hearing impaired listeners and, as a secondary goal, achieving appropriate loudness impressions for mid-level speech signals. The underlying model of hearing impairment assumes that the total hearing loss, HL, as quantified by audiometric thresholds, is comprised of a hearing-loss component attributable to outer hair cell (OHC) loss or damage, HL_{OHC}, and a component attributable to inner hair cell (IHC) loss or dysfunction (including damage of subsequent stages). It is assumed that HL_{OHC} manifests as low-level gain loss (GL) in the non-linear basilar membrane (BM) input-output function. Increasing HL_{OHC} would thus reduce the low-level gain and the compressive region of the BM I/O function shifting the compression knee-point to higher input levels while maintaining the compression ratio (see Plack et al., 2004). The MDC2 algorithm is meant to replace the altered or absent function of the OHCs by applying the missing level-dependent gain in narrowband auditory filters. To a first approximation, gain application is instantaneous only limited by the bandwidth of the filter. One fundamental problem with such an approach trying to replace the biological function of the non-linear BM processing in narrowband auditory filters is that the hearing aid algorithm has to re-synthesize an output signal from the multi-channel representation which is subsequently re-analyzed by the BM processing of the hearing impaired listener. Inevitably this re-analysis will cause “smearing” of signal energy originally intended for a specific narrowband frequency region across a wider frequency region. Thus gain applied to one auditory filter in the algorithm will affect a wider frequency region, violating the original aim to restore the frequency-specific OHC function. While this problem is not unique for model-based compression (see Kates, 1991), it becomes particularly apparent with the narrowband filter bands used in MDC2. For wideband signals, the aim to restore the OHC function can lead to effectively too high gains and loudness ratings, which are corrected for by a broadband control circuit in the MDC2 algorithm.

![Block diagram of the model-based dynamic compression algorithm MDC2.](image)

**FIGURE 1.** Block diagram of the model-based dynamic compression algorithm MDC2. Details are described in the text.
Figure 1 shows the block diagram of the MDC2 algorithm suggested here. The complex-valued, quasi-analytic signal at the output of a 4th-order Gammatone filterbank (30 bands, one ERB wide) was used to estimate the instantaneous level, $L_{\text{inst}}$, from the absolute value of the analytic signal, and the instantaneous frequency, $F_{\text{inst}}$, from the instantaneous phase of the analytic signal in each frequency band. In addition, a broad-band, “long-term” level $L_{\text{lt}}$ was estimated over five frequency bands centered around the respective frequency band using a smoothing time constant of 50 ms (1st-order lowpass). The maximum of a corrected $L_{\text{lt}}$ and the instantaneous level was used for further processing. A model for basilar membrane compression including off-frequency component suppression for normal-hearing (NH) and hearing-impaired (HI) subjects was computed in real-time, based on the estimated level and instantaneous frequency (Ewert and Grimm, 2011; Hohmann and Kollmeier, 2006). The difference between the modeled NH and individual HI BM-I/O function was applied as hearing aid gain per frequency band. The output signal of the algorithm was generated by a 2nd-order Gammatone re-synthesis filterbank including delay compensation between the channels (Hohmann, 2002).

MDC2 was fitted for individual hearing losses by estimation of $HLOHC$ or equivalently $GL$. Based on results of Jürgens et al. (2011), Ewert and Grimm (2011) suggested estimation of $GL$ from audiometric thresholds ($HL$) and the lower slope of the loudness function ($m_{\text{low}}$) measured with adaptive categorical loudness scaling (ACALOS; Brand and Hohmann, 2002):

$$GL = 0.58HL + 24.37m_{\text{low}} + 8.54.$$ (1)

The estimated $GL$ was used to modify the NH BM-I/O function to an individual HI BM-I/O function. The MDC2 algorithm compares the output of the NH and HI BM-I/O in real-time and derives the required gain from the difference. For details see Ewert and Grimm (2011).

**Evaluation of Aided Speech Perception Performance in Stationary and Fluctuating Noise**

Speech reception thresholds (SRT) were measured in stationary and fluctuating noise for three hearing loss compensation algorithms (Table 1 provides an overview of the parameters):

- **MDC2**: Compression algorithm MDC2 and fitting described above.
- **REF**: As a reference served a conventional multi-band compressor (Grimm et al., 2006) with nine $\frac{1}{4}$-octave frequency bands, and an attack-release filter with 20-ms attack and 100-ms release time of the level estimator in each band. The estimated input level was used for a lookup of level- and frequency dependent gains, as provided by the gain prescription. For gain prescription a gain table was calculated from the difference of the estimated NH and HI BM-I/O using identical model functions and fitting as described for MDC2.
- **LIN**: For a linear hearing loss compensation the prescribed gains for a 65-dB speech-shaped noise have been estimated in nine $\frac{1}{4}$-octave frequency bands and were applied independently of the input level in each of the frequency bands.

**TABLE 1.** Comparison of parameters of the three hearing loss compensation algorithms used in the speech intelligibility test. The bracketed values for the time constants in MDC2 are for the long-term level estimate $L_{\text{lt}}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MDC2</th>
<th>REF</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter bands</td>
<td>30</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Band width</td>
<td>1 ERB</td>
<td>$\frac{1}{4}$ octave</td>
<td>$\frac{1}{4}$ octave</td>
</tr>
<tr>
<td>Attack time constant</td>
<td>0 ms (50 ms)</td>
<td>20 ms</td>
<td>n/a</td>
</tr>
<tr>
<td>Decay time constant</td>
<td>0 ms (50 ms)</td>
<td>100 ms</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**Subjects, Stimuli, and Procedure**

16 hearing impaired subjects participated with moderate sensorineural hearing loss (age ranging from 45 to 77 years, median age 66 years; pure-tone average hearing loss, PTA, at 1, 2, 4 kHz of 40 dB HL, standard deviation 4 dB HL). Speech reception thresholds were measured adaptively using the Oldenburg sentence test (Wagener and Brand, 2005) in speech shaped stationary noise (SSN; same long-term spectrum as the sentences of the sentence test) and fluctuating background noise. Fluctuating background noise was either the international speech test signal (ISTS; Holube et al., 2010) or a mixture of two uncorrelated ISTSs (referred to as double ISTS in the following).
**Apparatus**

The speech intelligibility test was running on a standard PC. Target and noise signals were output by a RME ADI-8 Pro DA converter running at 48 kHz sampling rate. Further signal processing was implemented in a framework for low-delay real-time processing, the HoerTech Master Hearing Aid (MHA; Grimm et al., 2006). The electric output signal coming from the speech intelligibility test was sampled by a ESI Maya 44 USB sound card with 48 kHz sampling rate. Internally on a Linux based signal processing computer hosting the MHA, the signal was resampled to 16 kHz, processed, and upsampled to 48 kHz. One input channel was assigned to the target signal, and a second channel for level control of the background noise signal. In the MHA, virtual acoustics based on convolution with head-related impulse responses (HRIR; Kayser et al., 2009) was used. The target and noise signals were convolved with the HRIR for 0 degree azimuth. Switching of hearing loss compensation schemes was controlled via a network connection from the control computer, which also ran the speech test software. Signals were presented bilaterally via Sennheiser HD 650 headphones while the subjects were seated in a sound-insulating booth.

**Results and Discussion**

Figure 2 shows mean SRTs for the three noise conditions including ± one standard deviation ranges. Numbers indicate the individual subject’s data. Different gray shades represent the different hearing loss compensation schemes as indicated in the legend. It is obvious that SRTs are much improved in the single ISTS condition while they are comparable for the stationary speech-shaped noise (SSN) and the double ISTS conditions. Standard deviation across listeners is smallest for SSN and increases for ISTS and double ISTS. All three algorithms lead to comparable performance in for each of the noise conditions. As expected for NH listeners, SRTs are much improved in the fluctuating ISTS when compared to the stationary SSN condition. In this case, all three compensation schemes allow the HI listeners to take benefit of the fluctuating masker. In case of the double ISTS, temporal gaps are reduced and the overall modulation rate is increased. It appears plausible that performance is again more comparable to the SSN condition. Further evaluation of the suggested MDC2 compressor in bilateral and binaural configurations is presented in a companion paper (Ernst et al., 2013).

**FIGURE 2.** Speech reception thresholds (SRT) for the three noise conditions speech-shaped noise (SSN), ISTS and double ISTS. The black horizontal lines indicate average SRTs and the bars ± one standard deviation ranges. The three dynamic compressor schemes are represented by the different shades of gray. Numbers indicate individual data.
TOWARDS IMPROVED ESTIMATES OF OHC AND IHC LOSS OR DAMAGE

It was assumed that OHC loss or damage will alter the non-linear behavior of the BM-I/O function resulting in reduced low-level gain. The resulting GL was compensated by the MDC2 algorithm. Estimation of HL_{OHC} or GL was based on audiometric thresholds and loudness scaling. Following the assumption that the total hearing loss, HL, is comprised of a OHC- and IHC-related component (HL = HL_{OHC} + HL_{IHC}), any additional estimate of HL_{IHC} will improve the reliability of the HL_{OHC} estimate required to fit the MDC2 algorithm.

So far, IHC has often been simplified as a level-independent damping component, represented by reduced spike rates at the level on the auditory nerve. Consequently, IHC loss or damage has been implemented as linear damping in effective models of the impaired auditory system (e.g., Jepsen and Dau, 2011). The more physiological model by Zilany and Bruce (2006) incorporates IHC damage by changing the slope of the IHC transducer function (for a review of models for effects of hearing loss see Heinz, 2010). While reduced spike rates and reduced steepness of the rate-level function are major phenomena related to IHC damage, it can also be hypothesized that loss or dysfunction of IHC and subsequent stages will reduce the coding fidelity of temporal fine structure (TFS). In this case, not only spike rate would drop but also spike timing might be affected. Possible psychophysical consequences of such a reduced TFS coding fidelity are hypothesized to manifest in reduced performance in detection of low-rate frequency modulation (e.g., Strelcyk and Dau, 2009) or detection of frequency shifts in unresolved harmonic tone complexes (Moore et al., 2006). Moreover, Hopkins and Moore (2011) demonstrated age effects in TFS processing performance. If tests are run in background noise, interpretation of results is, however, further complicated: The typically widened auditory filters in HI listeners lead to decreased internal signal-to-noise ratios (SNR) at the output of the filter which in turn might degrade TFS coding acuity. A recent study on neural responses from noise-exposed fibers in chinchillas (Henry and Heinz, 2012) showed that peripheral temporal coding with sensorineural hearing loss is substantially more degraded in background noise than in quiet, which likely arises from the broader cochlear frequency tuning.

Here, detection thresholds for random frequency modulation in quiet and at two different SNRs were measured. The measurement is part of a series of five psychoacoustic experiments conducted in the same subjects and was chosen to exemplify the proposed approach of improving hearing loss estimates. The goal of the experimental series is to provide a wide range of near- and supra-threshold measurements as input to models of the impaired auditory system.

Detection of Random Frequency Modulation in Silence and Noise

**Subjects, Stimuli, and Procedure**

23 listeners participated in the random frequency modulation experiment. The listeners can be separated in three groups: Six normal-hearing young (NH-Y) listeners (25 to 31 years, median 27 years) with audiometric thresholds not worse than 20 dB HL up to 8 kHz. All NH-Y listeners were tested in the left ear. Six normal-hearing old (NH-O) listeners (aged between 65 and 77 years, median: 67) formed the second group. Their audiograms were “normal” (thresholds ≤ 20 dB HL) in the main speech region up to 4 kHz. Two of the six NH-O listeners had thresholds above 20 dB HL (35 and 40 dB HL) at 6 kHz. They were tested at their left ear or the better ear, if one ear was significantly better at higher frequencies. The third group consisted of eleven hearing-impaired (HI) old listeners (aged between 26 and 82 years, median 61). They all had a sensorineural, symmetric hearing loss, sloping to the higher frequencies. At the selected ear the mean hearing loss was 28 ±11, 51 ±13, 65 ±12 dB HL for the test frequencies of 500, 2000, and 6000 Hz, respectively. Seven of the HI listeners reported to use hearing aids in daily life.

The target stimuli were random frequency modulated (RFM) tones at 500, 2000, and 6000 Hz. The RFM tones were generated by adding a bandpass noise (2-4 Hz) as instantaneous frequency deviation to the tone’s frequency. The frequency modulation depth was expressed as root-mean-square (RMS) deviation of the instantaneous frequency from the center frequency (Δf_{RMS}/f). The reference stimuli were pure tones at 500, 2000, and 6000 Hz. Target and reference stimuli had an additional 2-4-Hz bandpass-noise amplitude modulation applied with a RMS modulation depth of -12 dB. The stimuli were 500 ms in duration separated by 300 ms pause intervals. RFM detection thresholds were measured in quiet and in the presence of a 5-ERB bandlimited Gaussian noise (ranging from 328-725, 1474-2689, and 4529-7926 Hz for center frequencies of 500, 2000, and 6000 Hz, respectively). Two noise levels were used yielding a SNR of +3dB and -3 dB.
An adaptive, 3-interval, 3-alternative, forced-choice procedure was used to estimate $\Delta f_{\text{RMS}}/f_c$ at threshold. The 70.7% point of the psychometric function was targeted using a 1-up, 2-down tracking rule. For each threshold run, the threshold estimate was calculated as the arithmetic mean over the last six reversals of the adaptive track. Five repetitions of each condition were measured; the geometric mean of the estimates from the last four repetitions was considered as final threshold estimate.

Stimuli were presented at 65 dB HL for the NH listeners. For the HI listeners, the level corresponding to medium loudness, as estimated from adaptive categorical loudness scaling, was used limited to a maximum level of 80 dB HL.

**Apparatus**

Stimuli were generated digitally in MATLAB with the AFC toolbox developed in Oldenburg, D/A-converted by a RME ADI-8 PRO or RME Fireface UC at 44.1 kHz sampling rate, and presented via a Tucker Davis HB7 headphone buffer and Sennheiser HDA 200 circumaural headphones in a double-walled, sound-attenuating booth. The headphones were calibrated with the Brüel&Kjær artificial ear (type 4153).

**Results and Discussion**

Figure 3 shows geometric mean and ± one standard deviation ranges of the RFM detection thresholds across all listeners per group as indicated by the different shades of gray (see legend). Numbers indicate data points for the individual listeners. A three-way analysis of variance (ANOVA) with factors subject group, center frequency and condition revealed that all factors were significant ($p<0.001$). A post hoc test with Fisher’s least-significant-difference (LSD) criterion showed that the means of group NH-O were significantly ($p>0.001$) higher than for the NH-Y group, most pronounced at 500 Hz for -3 dB SNR. Similarly, the means of the HI were higher than for the NH-O group. These differences decrease at 6 kHz, where a larger group overlap can be observed. The largest spreads inside a group are observed for the HI in clean condition. The age effect observed between the NH-Y and NH-O group is in line with results of Hopkins and Moore (2009).

**Towards Modelling Consequences of OHC and IHC Damage**

In order to better understand possible effects of OHC and IHC damage on the detection of RFM, the stimuli were passed through the auditory model of Meddis (2006). In the model, stimuli were filtered with a linear bandpass to model the middle-ear filtering and transformation in stapes velocity and then converted to BM velocity using the dual-resonance-non-linear (DRNL) filter (Meddis et al, 2001). This non-linear BM stage is suited to mimic OHC damage by reduction of gain in the non-linear pathway (see, e.g., Jepsen and Dau, 2011). BM velocity is then...
converted to membrane potentials in the IHC-stage of the model to finally generate probabilistic auditory nerve (AN) fiber spike responses including refractory effects and adaptation. In Fig. 4 (left) short segments of such generated spike trains from 10 high-spontaneous rate AN fibers all innervating the same location (500Hz) on the cochlear partition are shown. The spikes from one exemplary fiber are marked in red. The input stimuli were the reference signals from the RFM measurement (upper panel: 500-Hz pure tone in quiet, middle panel: same pure tone at +3dB SNR in 5-ERB-wide Gaussian noise, lower panel: same for -3dB SNR). The alignment of spikes to a specific phase of the tones (“phase-locking”) is clearly visible. The right panels show first-order inter-spike-interval histograms to demonstrate the distribution of the observed times between spikes averaged over all AN fibers. Phase-locking to the stimulus frequency of 500 Hz is represented by the maxima in the histogram which occur once each stimulus cycle (every 2 ms here). The strength of phase-locking to 500 Hz was quantified by the vector strength (VS) calculated for the 500-ms stimuli and averaged across 100 stimulus realizations. VS decreased from 0.63 to 0.43 with increasing noise level. The standard deviation of the VS is influenced by the amount of AN fibers, being higher for a reduced amount of fibers.

**FIGURE 4.** Simulated auditory nerve spike trains for 10 fibers (left) and inter-spike-interval histograms (right) constructed from the output of the auditory model by Meddis et al. (2006). Upper panels: Response to 500-Hz pure tone in quiet. Middle panels: Same pure tone at +3 dB SNR. Lower panels: Same for -3 dB SNR.

Random frequency modulation of the 500 Hz tone with a RMS modulation index of 5% decreases the average VS to 0.51, 0.45, and 0.35, respectively. If one assumes that RFM detection in the experiment is based on temporal coding for the 500-Hz stimuli, the difference of VS for RFM stimuli compared and the reference stimuli can be used to describe the detection performance. Detectability can be quantified by Cohen’s d defined here as the difference between the reference and target VS divided by their averaged standard deviation. Table 2 contains d’ for the three noise conditions from the experiment derived for a RMS modulation index of 5%. A NH subject and three principle types of hearing impairment were considered. The NH used the standard settings proposed by Meddis (2006) with 100 AN fibers for the one considered BM filter tuned to 500 Hz. The stimulus level was 65 dB HL (76 dB SPL) as in the experiment. IHC loss was modeled as a decrease of AN fibers from 100 to 10 (reducing the spike rate by a factor of 10, equivalent to 20 dB linear attenuation). This reduction of AN fibers also results in an increased standard deviation of the mean VS (reduced temporal coding acuity). OHC loss was realized by deactivating the non-linear path in the DRNL stage (equivalent to 38 dB gain loss) resulting in broader auditory filters which in turn results in a reduced VS for the noisy conditions (more noise energy falls into the filter). Combined IHC and OHC loss used both impairments at the same time. For the hearing impairment simulations the same stimulus level as for the NH was used. As expected, d’ decreases with increasing noise level for all four types of hearing. While IHC loss results in reduced d’ even without the presence of noise, OHC loss influences the detection process only for noisy conditions. Worst results were achieved for the combined impairment. This first model approach give rise to hope that IHC and OHC loss as well as temporal and AM cues for the detection of frequency modulation can be separated in the model.
and thus contribute to a better understanding of sensorineural hearing loss. In the empirical data HI (no noise) and NH-Y (+3 dB SNR) showed comparable average thresholds, which is also reflected by comparable d’ values (about 9) in Table 2 if IHC loss of combined hearing loss is assumed for the HI listener. However, d’ values were generally quite high. The model will be adapted to individual hearing loss in the future and threshold prediction will be derived for a criterion d’.

### TABLE 2. Estimated detectability (d’) for three different noise levels for NH and three types of impairments (IHC loss, OHC loss, combined IHC and OHC loss). The d’ was derived by dividing the difference of the mean vector strength for the reference signal and the RFM signal with 5-% RMS modulation index by the standard deviation of the means. Mean vector strength and standard deviation were estimated from 100 stimulus realizations.

<table>
<thead>
<tr>
<th>Condition</th>
<th>NH</th>
<th>IHC loss</th>
<th>OHC loss</th>
<th>combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>no noise</td>
<td>32.6</td>
<td>9.2</td>
<td>27.5</td>
<td>9.1</td>
</tr>
<tr>
<td>+3 dB SNR</td>
<td>8.9</td>
<td>6.0</td>
<td>5.4</td>
<td>4.7</td>
</tr>
<tr>
<td>-3 dB SNR</td>
<td>2.2</td>
<td>2.3</td>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### SUMMARY AND DISCUSSION

A model-based hearing aid compression algorithm (MDC2) was suggested aiming to restore the NH BM-I/O function in HI listeners. MDC2 was fitted by estimation of the OHC-loss related component, HL\textsubscript{OHC}, of the total hearing loss, equivalent to cochlear gain loss, derived from audiometric thresholds and adaptive categorical loudness scaling. The compression algorithm was evaluated measuring SRTs in 16 HI listeners with sensorineural hearing loss in stationary and fluctuating noise. SRTs were comparable for MDC2, a conventional 9-band reference compressor, and linear amplification.

To investigate psychoacoustic measures for estimation of the IHC-related hearing loss component, HL\textsubscript{IHC}, RFM (random frequency modulation) detection thresholds in quiet and in background noise were collected for six young NH listeners, six older NH listeners, and eleven HI listeners. A first model approach was presented for the 500-Hz condition to assess the differential effects of HL\textsubscript{OHC} and HL\textsubscript{IHC} on the detectability of RFM, accounting for the general trends in the data. Model predictions clearly show that HL\textsubscript{OHC} does not decrease RFM detectability in the quiet condition as was observed for the HI group in the empirical data suggesting that their hearing impairment is not solely based on OHC dysfunction. HL\textsubscript{IHC}, modeled as a reduction of AN fibers, did cause decrease RFM detectability independent of SNR in the stimuli.

The current modeling approach is rather limited and only assessed temporal phase locking at 500 Hz with vector strength as a measure. Moreover, just a single BM site (peripheral channel) tuned to the signal frequency was considered. The coding of temporal envelope fluctuations in the spike trains and multi-channel or across-channel analysis were not taken into account. A future model version will have to incorporate these cues to account for detection of RFM at all tested frequencies (500, 2000, and 6000 Hz). The role of purely temporal mechanisms as assumed so far appears less likely with increasing stimulus frequency.

### ACKNOWLEDGMENTS

We would like to thank Giso Grimm, Stephan Ernst, and Birger Kollmeier for fruitful contributions and discussions. We thank Jürgen Kießling and Hartmut Meister for their contributions in the SRT data collection. This work was funded by funded by BMBF 01EZ0741, DFG FOR1732 and DFG cluster of excellence Hearing for All.

### REFERENCES


