

# Spectral and temporal processing in the human auditory system

TORSTEN DAU<sup>1</sup>, MORTEN L. JEPSEN<sup>1</sup>, AND STEPHAN D. EWERT<sup>2</sup>

<sup>1</sup>*Centre for Applied Hearing Research, Ørsted•DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark*

<sup>2</sup>*Medical Physics, University of Oldenburg, D-26111 Oldenburg, Germany*

An auditory signal processing model is presented that simulates psychoacoustical data from a large variety of experimental conditions related to spectral and temporal masking. The model is based on the modulation filterbank model by Dau *et al.* [J. Acoust. Soc. Am. 102, 2892-2905 (1997)] but includes the dual-resonance non-linear (DRNL) filterbank suggested by Lopez-Poveda and Meddis [J. Acoust. Soc. Am. 110, 3107-3118 (2001)] to simulate the non-linear cochlear signal processing, as well as several other modifications at later processing stages motivated by other recent findings. The model was tested in conditions of tone-in-noise masking, intensity discrimination, spectral masking with tones and narrowband noises, forward masking with (on- and off-frequency) noise- and pure-tone maskers, and amplitude modulation detection using different noise carrier bandwidths. One of the key properties of the model is the combination of the fast-acting cochlear compression with the slower compression realized in the adaptation stage of the model. Both play a crucial role for the success of this model.

## INTRODUCTION

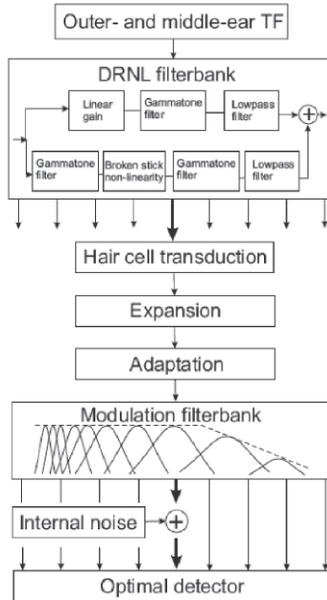
The perception model presented in Dau *et al.* (1997) was designed to account for human signal detection data in various psychoacoustic conditions. Rather than trying to model physiological details of auditory processing, the approach was to focus on the “effective” signal processing in the auditory system, which uses as little physiological assumptions and physical parameters as necessary, but tries to predict as many perceptual data as possible. The model has proven successful in predicting data from spectral and spectro-temporal masking (e.g., Verhey *et al.*, 1999; Derleth and Dau, 2000), non-simultaneous masking and modulation detection (Dau *et al.*, 1996, 1997; Ewert and Dau, 2004). In addition, for example, the preprocessing of the model has been used in objective assessment of speech quality (Hansen and Kollmeier, 1999).

However, the original model uses the gammatone filterbank to simulate peripheral filtering and thus does not include nonlinearities associated with basilar-membrane (BM) processing (e.g., Ruggero *et al.*, 1997). It can thus be expected that the model fails in conditions which reflect the nonlinear processing in the cochlea, such as forward masking with on- and off-frequency maskers (e.g., Oxenham and Plack, 2000) and spectral masking patterns as a function of the masker level (e.g. Moore *et al.*, 1998). Meddis *et al.* (2001) developed a non-linear cochlear model, the dual-resonance non-linear (DRNL) filterbank. They showed that their model can account for

several important properties of BM processing, such as frequency- and level-dependent compression and frequency selectivity. The DRNL structure and parameters were later adopted to develop a human cochlear filterbank model (Lopez-Poveda and Meddis, 2001) based on pulsation threshold data. In the present study, the linear gammatone filterbank stage in the original perception model (Dau *et al.*, 1997) was replaced by the DRNL filterbank. Some additional changes were undertaken in the subsequent stages of the overall model, motivated by recent findings mainly from studies on modulation perception (e.g., Ewert and Dau, 2000; Kohlrausch *et al.*, 2000). In the present study, the new model was tested in critical tasks of temporal and spectral masking.

**THE MODEL**

The new model (Figure 1) has a similar overall structure as the original model. The first stage is the DRNL filterbank (Lopez-Poveda and Meddis, 2001). The transformation of the mechanical BM oscillations into inner hair-cell receptor potentials is simulated roughly by half-wave rectification and low-pass filtering at 1-kHz. The signal is then transformed into an intensity-like representation, by applying a squaring expansion. This step is motivated by the findings in Müller *et al.* (1991) showing that the auditory-nerve spike rate as a function of stimulus level exhibits a square-law behaviour.



**Fig. 1:** Sketch of the auditory processing model. The model includes outer- and middle ear filtering, DRNL filtering on the BM, hair-cell transformation, expansion, adaptation, a modulation filterbank and an optimal detector as decision stage.

The adaptation stage in the model simulates adaptive properties of the auditory periphery. As in the original model, the effect of adaptation is realized by a chain of five feedback loops in series with different time constants. The output of the entire stage

approaches a logarithmic compression for stationary signals. For input variations that are rapid, compared with the time constants of the low-pass filters, the transformation through the adaptation loops is more linear, leading to a higher sensitivity for fast temporal variations. The output of the adaptation stage is filtered by a 1<sup>st</sup>-order low-pass filter at 150 Hz, motivated by results from modulation detection data with sinusoidal carriers. (e.g., Ewert and Dau, 2000; Kohlrausch *et al.*, 2000).

The low-pass filter is followed by a modulation filterbank as proposed in Dau *et al.* (1997). The lowest modulation filter is a 2<sup>nd</sup>-order lowpass filter with a cutoff frequency at 2.5 Hz. The modulation filters tuned to 5 and 10 Hz have a constant bandwidth of 5 Hz. For modulation frequencies at and above 10 Hz, the modulation filter center frequencies are logarithmically scaled and the filters have a constant Q value of 2. The magnitude transfer functions of the filters overlap at their -3 dB points. As in the original model, the modulation filters are complex frequency-shifted first-order lowpass filters. These filters have a complex valued output and either the absolute value of the output or the real part can be considered. For the filters centered above 10 Hz, the absolute value is considered. This is comparable to the Hilbert envelope of the bandpass filtered output and only conveys information about the presence of modulation energy in the respective modulation band, i.e., the modulation phase information is strongly reduced. This is in line with the observation of decreasing monaural phase discrimination sensitivity for modulation frequencies above about 10 Hz (Dau *et al.*, 1996; Thompson and Dau, 2008). For modulation filters centered at and below 10 Hz, the real part of the filter output is considered. In contrast to the original model, the output of these low-frequency modulation filters is multiplied by a factor of  $\sqrt{2}$ , so that the rms value at the output is the same as for the higher-frequency channels in response to a sinusoidal AM input signal of the same modulation depth.

Internal noise is added in order to limit the resolution of the model. The decision device is realized as an optimal detector (Dau *et al.*, 1996, 1997). The model was calibrated by adjusting the variance of the internal noise such that the model satisfies Weber's law when considering an intensity discrimination task using pure-tone stimuli.

## EXPERIMENTS

The model was tested in a variety of experimental conditions, including tone-in-noise simultaneous masking, forward masking, and modulation detection and masking (Jepsen *et al.* (2008)). The present papers focuses on the model's capabilities of predicting spectral masking and forward masking. The data for the spectral masking experiments were taken from Moore *et al.* (1998). The forward masking data represent own results (Jepsen *et al.*, 2008).

### Stimuli and procedure

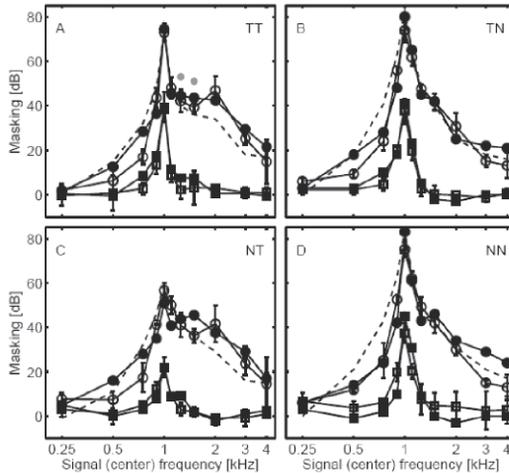
In the spectral masking experiment, the signal and the masker were either a pure tone or a 80-Hz wide Gaussian noise (Moore *et al.*, 1998). All four signal-masker configurations were considered: tone signal and tone masker (TT), tone signal and noise masker (TN), noise signal and tone masker (NT), and noise signal and noise masker (NN).

In the TT-condition, a 90-degree phase-shift between signal and masker was chosen, while the other conditions used random onset phases. The masker frequency was centred at 1 kHz, and the signal frequencies were 0.25, 0.5, 0.9, 1.0, 1.1, 2.0, 3.0 and 4.0 kHz. The signal and the masker were presented simultaneously. Both had a duration of 220 ms with 10 ms onset and offset squared-cosine ramps. Two masker levels were used: 45 and 85 dB SPL.

In the forward masking experiment, tonal signals and maskers were used. The stimuli were similar to those used in Oxenham and Plack (2000). Two conditions were considered: in the on-frequency masking condition, the signal and the masker were presented at 4 kHz. In the off-frequency condition, the signal frequency was still at 4 kHz whereas the masker frequency was 2.4 kHz. The signal had a duration of 10 ms and a Hanning window was applied to the entire signal duration. The masker was 200-ms long and had 2-ms ramps at the onset and the offset. The signal and the masker had random onset phases in both conditions. The signal level was varied during the experimental procedure and the signal level at masked threshold was obtained for a given masker level. In the on-frequency masking condition, the masker was presented at levels from 30 to 80 dB SPL, in 10-dB steps. In the off-frequency masking condition, the masker was presented at 60, 70, 80 and 85 dB SPL. The separation between masker offset and signal onset was either 0 ms or 30 ms.

## RESULTS

### Spectral masking patterns



**Fig. 2:** Spectral masking patterns from the stimulus conditions TT, TN, NT and NN. Squares and circles show results for a masker level of 45 and 85 dB SPL, respectively. Open symbols indicate data (from Moore *et al.*, 1998) while closed symbols represent simulations. The dashed curve shows simulation obtained with linear BM processing (from Derleth and Dau, 2000).

Spectral masking patterns are plots of the amount of masking of a signal as a function of the signal frequency in the presence of a masker (with fixed frequency and level). The shapes of these masking patterns are influenced by a variety of factors, such as occurrence of combination tones or harmonics (produced by the peripheral non-linearities), beating cues, and resolved spectral components. The mean data from Moore *et al.* (1998) for the 1-kHz masker are shown in Fig. 2 as open symbols. The simulated masking patterns obtained with the current model are indicated by the filled symbols. In addition, simulations using the original processing model are shown by the dashed curves (Derleth and Dau, 2000).

Panels A to D show the results for the different signal-masker conditions (TT, TN, NT, NN). Two masker levels are considered in each configuration: 45 dB SPL (squares) and 85 dB SPL (circles). The ordinate represents masking, defined as the difference between the masked signal threshold and the corresponding signal threshold in quiet. The masking patterns in the four conditions generally show a maximum at the masker frequency. The amount of masking decreases with increasing spectral separation between the signal and the masker. The 45-dB SPL masker produces a symmetric pattern in all conditions, whereas the pattern for the 85-dB masker is asymmetric with a broadening towards higher frequencies.

For the TT condition (panel A), the amount of tuning in the masking patterns is particularly strong since beating between the signal and the masker provides a very effective detection cue in this condition. The predictions agree well with the experimental data, except for the threshold for the signal frequencies 500 and 750 Hz for the high masker level (85 dB), where the amount of masking is overestimated. The gray circles show additional simulations where only the first 8 modulation filters were included (with center frequencies from 0 to 130 Hz) whereas modulation channels tuned to higher frequencies were not considered. These additional predictions clearly overestimate the amount of masking, suggesting that beating between the signal and the masker with rates of 150 - 200 Hz provides an effective cue in this masking condition.

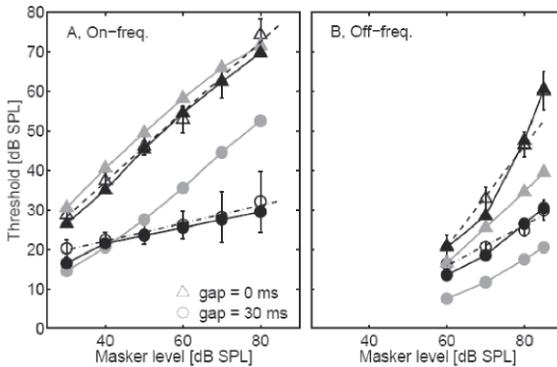
For the tonal signal and noise masker (TN, panel B), the masking pattern is broader than in the TT-condition at frequencies close to the masker frequency; the strong peak at 2 kHz was not observed for the noise masker. This is also reflected in the simulations. On the low-frequency side of the masker, the predictions are considerably better than those obtained with the original model. Thus, in this condition where energy cues play the most important role, the shapes of the level-dependent BM filters are mainly responsible for the good agreement between data and simulations.

For the NT condition (panel C), the amount of masking for the on-frequency situation is about 20 dB lower than in the previous two conditions (TT, TN). The reason for this “asymmetry of masking” effect is that signal detection for this on-frequency condition is based on the temporal structure of the stimuli (and not on energy), when the signal bandwidth is greater than the masker bandwidth (Hall, 1997). The simulated patterns agree well with the measured data, except for the signal frequencies 500 and 750 Hz at 85 dB SPL masker level.

Finally, the masking patterns in the NN-condition (panel D) are similar to those of the TN-condition. The simulations agree well with the measured patterns while the results obtained with the original model (dashed curve) clearly overestimate the masking on the low-frequency side of the masker by up to about 20 dB.

### Forward masking with on- versus off-frequency tone maskers

The forward masking experiment of the present study was considered in order to test the ability of the new model to account for data that have previously been explained in terms of the nonlinear BM processing (e.g., Oxenham and Plack, 2000). It was shown that if masker and signal level (in the on-frequency condition) lie within the compressive region of BM input/output function, the signal level at threshold changes linearly with changing masker level, i.e. reflecting a linear growth of masking (GOM) function. This is typically the case for very short masker-signal separations. In contrast, for larger temporal masker-signal separations, when the masker level may fall in the compressive and the signal level in the linear region of the BM input-output function, a change in masker level will produce a smaller change of the signal level at threshold. This causes a shallower slope of the GOM function. For off-frequency stimulation, with a masker frequency well below the signal frequency, the BM response at the signal frequency is assumed to be linear at all levels. The slope of the curves should therefore be roughly independent of the masker-signal separation for off-frequency stimulation.



**Fig. 3.** GOM functions from the forward masking experiment. Panel A and B show the on-frequency and off-frequency condition, respectively. Triangles indicate a gap of 0 ms and circles a gap of 30 ms. Open symbols indicate data while black and gray symbols represent simulations with non-linear and linear BM processing, respectively.

Figure 3 shows the measured data from the own experiment, averaged across four subjects. Signal level at threshold is shown as a function of the masker level, reflecting GOM curves. The left and right panels show the results for the on-frequency and off-frequency conditions, respectively. Thresholds corresponding to a masker-signal separation of 0 ms are indicated by triangles, and circles show the results for a masker-signal separation of 30 ms. In the on-frequency condition (left panel), the measured GOM

function is close to linear (0.9 dB/dB) for the 0-ms separation. For the larger masker-signal separation of 30 ms, the slope of the growth of masking function is more compressive (0.25 dB/dB) since signal and masker can be assumed to be processed in different level regions of the BM input-output function. The data agree with the results from Oxenham and Plack (2000) in terms of the slope of the GOM functions (0.82 dB/dB for the 0-ms gap, and 0.29 dB/dB for the 30-ms gap).

The corresponding simulations are shown as filled symbols in the same figure. The simulated GOM functions for both masker-signal separations are close to the measured data. This supports the hypothesis that the non-linear BM stage can account for the different shapes of the forward masking conditions observed for different separations. For direct comparison, simulations obtained with the original model (Dau *et al.*, 1997), using a gammatone filterbank, are represented by the filled gray symbols. Since this BM stage processes sound linearly, the slopes of the GOM functions are similar for the two masker-signal separations, in contrast to the data.

The right panel of Fig. 3 shows the results for the off-frequency condition. The data (open symbols) show a 1.2 dB/dB slope of the GOM function for the 0-ms masker-signal separation, and a 0.5 dB/dB slope for the 30-ms separation. These data are not in line with the hypothesis that the GOM function for off-frequency stimulation should be independent of the gap-size. The data also differ from the average data in Oxenham and Plack (2000) who found GOM functions in this condition with a slope close to one for all masker-signal separations. However, their average data showed substantial variability; some of their individual subject's data were clearly compressive while others were linear or slightly expansive. The corresponding simulations of the off-frequency condition are represented by the filled symbols. The simulations agree well with the measured data from the present study. Within the model, the slightly compressive GOM functions are caused by the adaptation stage, which compresses the long-duration off-frequency masker slightly more than the short-duration signal. This slight compression can thus also be seen in the simulations obtained with the original model (gray symbols). In the 0-ms condition, the signal threshold levels lie generally in the compressive part (>30 dB SPL) of the BM input/output function. As a consequence, the GOM function is less compressive since the masker is still processed linearly.

## DISCUSSION

Several major modifications were introduced into the original perception model (Dau *et al.*, 1997). The linear peripheral filterbank was replaced by the DRNL filterbank in order to account for the nonlinear processing at the level of the BM. Several additional changes such as a squaring expansion and modifications in the processing of amplitude modulation were introduced, motivated by findings from other recent modeling studies. The question was to what extent the new model would be able to keep (and extend) the capabilities of the original model of predicting results from various perceptual data. Here, spectral masking patterns and forward masking were considered. In the spectral masking task, signal detection is typically based on intensity cues, beating cues or resolved spectral components, depending on the specific signal-masker

configuration. These masking patterns are therefore interesting (and challenging) to test for any perception model. In the framework of the present model, the data can be accounted for by the combination of a (close to) logarithmic overall compression of the stimuli (realized mainly in the adaptation stage) with a high sensitivity to beatings between frequency components (realized in the modulation filterbank) and a “realistic” stage of peripheral frequency selectivity (realized in the DRNL).

As a possible explanation for forward masking, mainly two different mechanisms have been discussed in the past: (i) Persistence of neural activity (e.g., Oxenham and Moore, 1994), referring to temporal integration of neural activity at presumably higher stage the auditory nerve; and (ii) neural adaptation (e.g., Nelson and Swain, 1996) assuming adaptation at various levels of the auditory pathway. The temporal window model (e.g., Oxenham and Moore, 1994) represents a temporal-integration mechanism while the model of the current study represents an adaptation mechanism. The temporal window model was shown to account for the on-frequency and off-frequency forward masking data (e.g., Oxenham, 2001) in normal-hearing and hearing-impaired listeners. However, it should be noted that the decision mechanism in the temporal window model is based on the signal-to-masker (S/N) ratio at the output. It has been shown recently that the combination of integration and S/N detection criterion in the model acts essentially as adaptation (Ewert *et al.*, 2006). The adaptation model might be the more general approach since it shows the effect of adaptation in the internal representation of the stimuli, similar to that observed in neural responses, and can be applied successfully to probably a broader class of experimental masking conditions than the temporal window model. Thus, the combination of fast-acting BM compression, followed by fast acting (neural) expansion and a slower logarithmic compression allows the model to account for intensity-discrimination (Jepsen *et al.*, 2008) and simultaneous masking as well as forward masking.

Shamma and colleagues (e.g., Chi *et al.*, 1999; Elhilali *et al.*, 2003) described a model that is conceptually similar to the present model but includes an additional “dimension” in the signal analysis. They suggested a spectro-temporal analysis of the envelope, motivated by neurophysiological findings in the auditory cortex (Schreiner and Calhoun, 1995; deCharms *et al.*, 1998). In their model, a “spectral” modulation filterbank was combined with the temporal modulation analysis, resulting in 2-dimensional spectro-temporal filters. Thus, in contrast to the implementation presented here, their model contains joint (and inseparable) spectral-temporal modulations. In conditions where both temporal and spectral features of the input are manipulated, the two models respond differently. The model of Shamma and co-workers has been utilized to account for spectro-temporal modulation transfer functions, for the assessment of speech intelligibility (Chi *et al.*, 1999; Elhilali *et al.*, 2003), the prediction of musical timbre (Ru and Shamma, 1997), and the perception of certain complex sounds (Carlyon and Shamma, 2003). The present model is sensitive to spectral envelope modulation which is reflected as a variation of the energy (considered at the output of the modulation lowpass filter) as a function of the audio-frequency (peripheral) channel. For temporal modulation frequencies below 10 Hz, where the phase of the envelope

lope is preserved, the present model could thus use spectro-temporal modulations as a detection cue. The main difference to the model of Chi *et al.* (1999), however, is that the present model does not include joint spectro-temporal channels. It is not clear to the authors of the present study to what extent detection or masking experiments can assess the existence of joint spectro-temporal modulation filters. The assumption of the model presented here that (temporal) modulations are processed independently at the output of each auditory filter, implies that no across-channel modulation processing can be accounted for. This reflects a limitation of this model.

Recently, comodulation masking release (CMR) has been modeled using an equalization-cancellation (EC) mechanism for the processing of activity across audio frequencies (Piechowiak *et al.*, 2007). The EC process was assumed to take place at the output of the modulation filterbank for each audio-frequency channel. In that model, linear BM filtering was assumed. The model developed in the present study will allow a quantitative investigation of the effects of nonlinear BM processing, specifically the influence of level-dependent frequency selectivity, compression and suppression on CMR. The model might be valuable when simulating the numerous experimental data that have been described in the literature, and might in particular help interpreting the role of within-versus across-channel contributions to CMR. Another challenge will be to extend the model to binaural processing. The model of Breebaart *et al.* (2001) accounted for certain effects of binaural signal detection, while their monaural pre-processing was based on the model of Dau *et al.* (1996), i.e., without BM nonlinearity and without the assumption of a modulation filterbank. Effects of BM compression (Breebaart *et al.*, 2001) and the role of modulation frequency selectivity (Thompson and Dau, 2008) in binaural detection have been discussed, but not yet considered in a common modeling framework.

An important perspective of this model is the simulation of hearing loss and its consequences for perception. This may be possible because the model now includes realistic cochlear compression and level-dependent cochlear tuning. Cochlear hearing loss is often associated with lost or reduced compression (Moore, 1995). Lopez-Poveda and Meddis (2001) suggested how to reduce the amount of compression in the DRNL to simulate a loss of outer hair-cells for moderate and severe hearing loss. This could be used in the present modeling framework as a basis to predict the outcome of a large variety of psychoacoustic tasks in (sensorineural) hearing-impaired listeners.

## REFERENCES

- Breebaart, J., van de Par, S., and Kohlrausch, A. (2001a). "Binaural processing model based on contralateral inhibition. I. Model structure.," *J. Acoust. Soc. Am.* **110**, 1074-1088.
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). "Spectro-temporal modulation transfer functions and speech intelligibility," *J. Acoust. Soc. Am.* **106**, 2719-2732.
- Carlyon, R. P., and Shamma, S. (2003). "An account of monaural phase sensitivity," *J. Acoust. Soc. Am.* **114**, 333-348.

- de Charms, R. C., Blake, D. T., and Merzenich, M. M. (1998). "Optimizing sound features for cortical neurons," *Science* **280**(5368), 1439-1443.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow band carrier," *J. Acoust. Soc. Am.*, **102**, 2892-2905.
- Dau, T., Püschel, D., and Kohlrausch (1996). "A quantitative model of the effective signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am.* **99**, 3615-3622.
- Derleth, R. P., and Dau, T. (2000). "On the role of envelope fluctuation processing in spectral masking," *J. Acoust. Soc. Am.* **108**, 285-296.
- Elhilali, M., Chi, T., and Shamma, S. (2003). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," *Speech Commun.* **41**, 331-348.
- Ewert, S. D., and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* **108**, 1181-1196.
- Ewert, S. D., and Dau, T. (2004). "External and internal limitations in amplitude-modulation processing," *J. Acoust. Soc. Am.* **116**, 478-490.
- Ewert, S. D., Hau, O., and Dau, T. (2006). "Forward masking: temporal integration or adaptation?," in *Hearing – from basic research to applications.*, International symposium on Hearing, edited by Birger Kollmeier *et al.*, 165-174.
- Hall, J., (1997) Asymmetry of masking revisited: generalization of masker and probe bandwidth. *J. Acoust. Soc. Am.* **101**, 1023-1033.
- Hansen, M., and Kollmeier, B. (1999). "Continuous assessment of time-varying speech quality," *J. Acoust. Soc. Am.* **106**, 2888-2899.
- Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). "A computational model of human auditory signal processing and perception", *J. Acoust. Soc. Am.* (2008). Accepted.
- Kohlrausch, A., Fassel, R., and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," *J. Acoust. Soc. Am.* **108**, 723-734.
- Lopez-Poveda, E. and Meddis, R. (2001). "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Am.* **110**, 3107-3118.
- Meddis, R., O'Mard, L.P., and Lopez-Poveda, E.A. (2001). "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.* **109** (2001) 2852-2861.
- Moore B. C. J. (1995). *Perceptual Consequences of Cochlear Damage*. Oxford University Press, New York.
- Moore, B. C. J., and Alcántara, J. I. (1998): Masking patterns for sinusoidal and narrow-band noise maskers. *J. Acoust. Soc. Am.* **104**, 1023-1038.
- Muller, M., Robertson D., and Yates, G. K. (1991). "Rate-versus-level functions of primary auditory nerve fibres: evidence for square law behaviour of all fibre categories in the guinea pig," *Hearing Research* **55**, 50-56.
- Nelson, D. A., and Swain, A. C. (1996). "Temporal resolution within the upper accessory excitation of a masker," *Acta Acustica* **82**, 328-334.
- Oxenham, A. J., and Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking," *Hearing Research* **80**, 105-118.

- Oxenham, A. J., and Plack, C. J. (2000): Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity. *Hearing Research* **150**, 258-266.
- Oxenham, A. J. (2001). "Forward masking: Adaptation or integration?," *J. Acoust. Soc. Am.* **109**, 732-741.
- Piechowiak, T., Ewert, S. D., and Dau T. (2007). "Modeling comodulation masking release using an equalization-cancellation mechanism," *J. Acoust. Soc. Am.* **121**, 2111-2126.
- Ru, P., and Shamma, S. A. (1997). "Representation of musical timbre in the auditory cortex," *J. of New Music Res.* **26**, 154-169.
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* **101**, 2151-2163.
- Schreiner, C. E., and Calhoun, B. (1995). "Spectral envelope coding in cat primary auditory cortex: Properties of ripple transfer functions," *J. Auditory Neuroscience* **1**, 39-61.
- Thompson, E., and Dau, T. (2008). "Frequency selectivity in binaural processing of fluctuations in interaural level difference," *J. Acoust. Soc. Am.* **123**, 1017-1029.
- Verhey, J. L., Dau, T., and Kollmeier, B. (1999). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," *J. Acoust. Soc. Am.* **106**, 2733-2745.

