Spectro-temporal analysis of complex sounds in the human auditory system

Ph.D. thesis by

Tobias Piechowiak

CAHR
Centre for Applied Hearing Research

Technical University of Denmark
2009

ISBN 978-87-92465-13-9
Printed in Denmark by Hillerød Grafisk.

Abstract

Most sounds encountered in our everyday life carry information in terms of temporal variations of their envelopes. These envelope variations, or amplitude modulations, shape the basic building blocks for speech, music, and other complex sounds. Often a mixture of such sounds occurs in natural acoustic scenes, with each of the sounds having its own characteristic pattern of amplitude modulations. Complex sounds, such as speech, share the same amplitude modulations across a wide range of frequencies. This "comodulation" is an important characteristic of these sounds since it can enhance their audibility when embedded in similar background interferers, a phenomenon referred to as comodulation masking release (CMR). Knowledge of the auditory processing of amplitude modulations provides therefore crucial information for a better understanding of how the auditory system analyses acoustic scenes.

The purpose of the present thesis is to develop a computational auditory processing model that accounts for a large variety of experimental data on CMR, in order to obtain a more thorough understanding of the basic processing principles underlying the processing of across-frequency modulations.

The second chapter introduces a processing stage, in which information from different peripheral frequency channels is combined. This so-called across-channel processing is assumed to take place at the output of a modulation filterbank, and is crucial in order to account for CMR conditions where the frequency spacing of comodulated components is relatively large.

The third chapter investigates the role of nonlinear inner-ear (cochlear) processing on CMR. A compressive non-linearity is incorporated in the modeling framework suggested in the second chapter. This non-linearity is necessary to account for CMR in conditions which are sensitive to cochlear suppression.

The fourth chapter examines the role of cognitive processing in different stimulus

paradigms: CMR, binaural masking level differences and modulation detection interference are investigated in contexts of auditory grouping. It is shown that auditory grouping can influence the results in conditions where the processing in the auditory system is dominated by across-channel comparisons.

Overall, this thesis provides insights into the specific mechanisms involved in the perception of comodulated sounds. The results are important as a basis for future models of complex modulation processing in the human auditory system.

Resumé

De fleste lyde, som vi udsættes for i vores hverdag, indeholder information i form af tidsmæssige variationer i deres indhyllingskurve. Variationerne i indhyllingskurven eller amplitudemodulationerne udgør grundelementerne i tale, musik og andre former for sammensatte eller komplekse lyde. Naturligt forekomne akustiske situationer består ofte af en blanding af forskellige lydkilder med hver deres karakteristiske sammensætning af amplitudemodulationer (såkaldte amplitudemodulations mønster). Komplekse lyde som f.eks. tale indeholder de samme amplitudemodulationer i et stort frekvensområde. Denne "comodulation" er en vigtig egenskab, eftersom den kan forstærke hørbarheden af sådanne lyde i situationer, hvor de omgives af forstyrrende lyde af samme karakter. Dette fænomen kaldes for "comodulation masking release" (CMR). Kendskab til bearbejdningen/processeringen af amplitude-modulationer i den menneskelige hørelse giver derfor afgørende informationer i forbindelse med at opnå en bedre forståelse af, hvordan hørelsen analyserer akustiske situationer og omgivelser.

Formålet med denne afhandling er at udvikle en beregningsmæssig model af den databehandling/processering, der finder sted i hørelsen og som kan redegøre for en omfattende samling af forskellige eksperimentelle data. Dette vil medvirke til en mere indgående forståelse af de fundamentale principper, som ligger til grund for bearbejdningen/processeringen af amplitude-modulationer på tværs af forskellige frekvenskanaler.

Kapitel to præsenterer en databehandlingsblok, hvor information fra forskellige frekvenskanaler bliver integreret. Integrationen antages at finde sted efter, at lyden er blevet bearbejdet/processeret af en modulationsfilterbank, og den er afgørende for at kunne redegøre for CMR i situationer, hvor frekvensafstanden mellem de sammenhørige modulationer er relativt stor.

I kapitel tre undersøges sammenhængen mellem CMR og den ulineære behandling af lyden, der finder sted i det indre øre. En ulineær kompression er inkluderet i modellen, som er blevet introduceret i kapitel 2. Dette er nødvendigt, for at modellen kan redegøre for omstændigheder som (er følsomme over for/påvirkes af) den ulineære kompression i det indre øre.

Kapitel fire undersøger betydningen af kognitive processor i forbindelse med forskellige typer af stimulus: CMR, "binaural masking level differences" og "modulation detection interference" undersøges i sammenhæng med grupperingen af lyde i høresystemet. Det vises, at grupperingen af lyde i høresystemet kan påvirke resultaterne i situationer, hvor processeringen i høresystemet domineres af sammenligninger på tværs af forskellige frekvenskanaler.

Overordnet giver denne afhandling indsigt i de specifikke mekanismer, der er involveret i opfattelsen af sammenhørigt modulerede lyde. Resultaterne danner et vigtigt grundlag for fremtidige modeller af den komplekse processering af modulationer i den menneskelige hørelse.

Preface

Finally...the last words typed, the last changes made... a really overwhelming feeling. The essence of four years of my life. I enjoyed them professionally as well as personally. At this point I want to thank...

My supervisor Torsten Dau. Thank you for your patience with me and for giving me the freedom to find my own way of doing this work. I learnt a lot from you!

Stephan D. Ewert, who co-supervised this work for his invaluable support and discussions concerning the computational modeling. It was a pleasure to work with you!

The software Skype® for making this sometimes long discussions possible.

The whole team at the Centre for Applied Hearing Research for the great "hyggelig" atmosphere and the personal as well as professional help. In particular, I would like to thank Morten L. Jepsen, Sebastien Santurette, Iris Arweiler, Eric R. Thompson and Claus F. Christiansen for sometimes long breathed hours of participation in my listening experiments.

The Danish Research Foundation, the Danish Graduate school SNAK "Sense organs, neural networks, behavior, and communication", and the danish state for its hospitality.

My family and girlfriend for their unconditional support...

Tobias Piechowiak, 10th July 2009

Contents

1

General Introduction

The auditory system performs a complex transformation of the sound energy incident at our ears into percepts which enable us to orient ourselves and other objects within our surroundings. One of the major aims of psychoacoustic research is to establish functional relationships between the basic physical attributes of sound, such as intensity, frequency and changes of these in these characteristics over time, and their associated percepts. The present study deals particularly with the dimension of time in auditory processing. With most sounds in our environment, such as speech and music, information is contained to a large extent in the changes of sound parameters with time, rather than in the stationary sound segments. Temporal processing and resolution typically refers to the processing of the envelope of a sound, i.e. its envelope variations or amplitude modulations, rather than the fine structure of a sound referring to the variations of instantaneous pressure.

Speech, music and animal vocalization are characterized by coherent amplitude modulations across a wide range of (audio) frequencies. The ability to process such information is thought to be a powerful survival strategy in the natural world, aiding in the detection of target sounds in the presence of competing sounds. A simple example for such a benefit is the phenomenon of comodulation masking release (CMR). In CMR, the audibility of a target sound embedded in another masking sound can be improved by adding sound energy that is remote in frequency from both the masker and the target (Hall et al., 1984). An improvement, i.e. a release from masking, is observed when the remote sound and the masker share coherent patterns of amplitude modulation.

Even though CMR has been investigated in many studies, the underlying mechanisms have not been clarified. It has been postulated earlier that part of the CMR effect results from so-called "across-channel" comparisons of temporal

envelopes (Buus, 1985) whereby across-channel refers to an operation that compares information at the output of different auditory filters, or channels, after processing of the incoming sound though the inner ear, the cochlea. However, it also has been proposed that so-called "within-channel" cues, i.e., information from only the one auditory filter tuned to the signal frequency, can account for a considerable part of the effect in some conditions (Schooneveldt and Moore, 1987). This conclusion was supported by quantitative predictions provided by Verhey et al. (1999) using an auditory model that considered only the processing in a single peripheral channel in such CMR experiments. Furthermore, some authors have proposed that certain aspects of nonlinear processing of sound through the cochlea, associated with compression of sound level, influence the amount of observed CMR. In order to account for such effects it has been suggested to include level-dependent nonlinear processing in the modeling (e.g., Ernst and Verhey, 2006). Finally, it has recently been demonstrated in several experimental studies (Grose and Hall, 1993; Dau et al., 2009), that the amount of CMR also depends on the acoustical context of the stimuli: depending on the sound stimulation prior to or subsequent to the masker components, CMR can be reduced or even eliminated. This gave rise to an interpretation that CMR needs to be interpreted in terms of auditory grouping effects.

While different processing principles and models have been suggested in the past to account for CMR (e.g. Buus, 1985; Schooneveldt and Moore, 1987; Verhey et al., 1999), most of the descriptions have either been at a rather qualitative level or have only focused on one particular aspect of CMR. The main goal of the present thesis has been to develop a computational auditory processing model that accounts for a large variety of experimental data on CMR. Here, the attempt has been to develop a framework that covers the results from many different experimental paradigms while keeping the model parameters constant. Another important aspect has been to provide a model that is consistent with earlier results on numerous other phenomena on detection, discrimination and masking, such that both the new conditions (on CMR) as well as the key findings from earlier investigations can be successfully described.

In Chapter 2 of this thesis, an across-channel processing stage is described that can account for CMR in experimental conditions where stimulus information is compared across large spectral distances such that within-channel processes only

play a minor role. The assumed across-channel mechanism is based on concepts of binaural, i.e. across-ear, processes that have been established in earlier studies on binaural masking (e.g. Durlach, 1963). While the across-ear processing assumed a comparison at an early stage directly following cochlear processing, the across-channel process in the CMR model presented here is assumed to take place at a more central level of processing. The across-channel modulation processing stage is validated in several critical experimental conditions.

Chapter 3 investigates the role of nonlinear cochlear processing on CMR. While the processing in the previous chapter was based on a linear model of cochlear filtering, here some of the nonlinear properties are accounted for by a so-called non-linear dual-resonance non-linear (DRNL) filter stage as recently suggested by (Meddis et al., 2001). The crucial part of this DRNL filter is a compressive non-linearity in one of the two parallel processing paths; the remaining part of the overall processing is otherwise left unchanged. In the framework of the model, the role of compression on CMR is investigated and evaluated in several experimental conditions. In particular, the effects of (absolute) masker level, masker-signal level ratios as well as the dependence of the spectral distance between masker and signal components are investigated, all reflecting conditions that challenge the nonlinear extension of the proposed model for CMR.

Chapter 4 investigates effects of auditory grouping on CMR. Sound components are provided after the offset of the remote masker components in such a way that they are perceptually grouped together with the masker components in a sequential stream. With such an arrangement, the masker components are perceptually segregated from the target which leads to an elimination of CMR if CMR is associated with auditory grouping. In order to investigated the principles of auditory grouping on across-channel modulation processing, also two other well known phenomena of modulation channel processing are investigated: modulation detection interference and the phenomenon of binaural masking release. The results from this chapter are expected to provide constraints for models of complex modulation processing.

Finally, Chapter 5 summarizes the main findings of this thesis, discusses links to recent developments in related areas of auditory modeling and provides suggestions for future investigations within auditory modeling and perception research.

2

Modeling comodulation masking release using an equalization-cancellation mechanism

This chapter [1] presents an auditory processing model that accounts for the perceptual phenomenon of comodulation masking release (CMR). The model includes an equalization-cancellation (EC) stage for the processing of activity across the audio-frequency axis. The EC process across frequency takes place at the output of a modulation filterbank assumed for each audio-frequency channel. The model was evaluated in three experimental conditions: (i) CMR with four widely spaced flanking bands in order to study pure across-channel processing, (ii) CMR with one flanking band varying in frequency in order to study the transition between conditions dominated by within-channel processing and those dominated by across-channel processing, and (iii) CMR obtained in the "classical" band-widening paradigm in order to study the role of across-channel processing in a condition which always includes within-channel processing. The simulations support the hypothesis that within-channel contributions to CMR can be as large as 15 dB. The across-channel process is robust but small (about 2-4 dB) and only observable at small masker bandwidths. Overall, the proposed model might provide an interesting framework for the analysis of fluctuating sounds in the auditory system.

[1] *This chapter was originally published as Piechowiak et al. (2007)*

2.1 Introduction

Many properties of auditory masking can be understood in terms of the responses of the basilar membrane within the inner ear. Each part of this membrane behaves like a filter that responds to a limited range of frequencies. When trying to detect a sinusoidal tone in background noise, it has been proposed that listeners use the output of a single auditory filter tuned to the frequency of the tone (Fletcher, 1940). That filter passes the tone at full intensity, but rejects most of the background noise. Although this theory can account for many aspects of masking, Hall et al. (1984) and others showed that, when comodulated maskers were used, some of the results can be explained only if it is assumed that stimulus information is processed across the outputs of auditory filters. In fact, humans are often much better at detecting signals in comodulated maskers than in white noise, an effect called comodulation masking release (CMR; Hall et al., 1984). Various experiments on CMR have demonstrated that the human auditory system can exploit coherent envelope fluctuations very effectively and that substantial reductions in signal threshold can result. Since coherent across-frequency modulation is common in speech, music, animal vocalization and environmental noise, the ability to process such information is thought to be a powerful survival strategy in the natural world which aids in the detection of target sounds in the presence of competing sounds.

CMR was demonstrated initially by Hall et al. (1984). In their "band-widening" experiment, the detection of a tone was measured as a function of the bandwidth of a noise masker, keeping the spectrum level constant. They used two types of maskers. One was a random noise with irregular fluctuations in amplitude that are independent in different frequency regions. The other was a random noise which was amplitude modulated using a low-pass filtered noise as a modulator. This modulation resulted in slow fluctuations in the amplitude of the noise that were the same in different frequency regions. For the random noise, the signal threshold increased as the masker bandwidth increased up to about the critical bandwidth at that frequency and then remained constant, as expected from the classical power spectrum model of masking (Fletcher, 1940; Patterson and Moore, 1986). The pattern for the modulated noise was quite different. Here, the threshold decreased as the bandwidth was increased

beyond about 100 Hz (for a signal frequency of 2 kHz); thus, adding more noise to the masker made the signal easier to detect. This suggested that subjects may compare the outputs of different auditory filters to enhance signal detection. The fact that the decrease in threshold with increasing bandwidth only occurred with the modulated noise indicated that fluctuations in the masker are critical and that the fluctuations need to be correlated across frequency bands.

In a second class of experiments, CMR was demonstrated by using narrow bands of noise (of typically 20-50 Hz width), which inherently have relatively slow amplitude fluctuations. One band, the on-frequency band, was centered at the signal frequency. A second band, the flanker band, was placed remote from the signal frequency. When the flanking band was uncorrelated with the on-frequency band, there was typically no effect on signal threshold. However, when the flanking band was correlated with the on-frequency band, a flanking band produced a release from masking (Hall et al., 1984; Schooneveldt and Moore, 1987; Cohen and Schubert, 1987). CMR was also found even if the signal and on-frequency band were presented to one ear and the flanking band to the other ear (Schooneveldt and Moore, 1987; Cohen and Schubert, 1987).

Even though CMR has been investigated in a number of studies, the underlying mechanisms are still not clear. It has generally been assumed that CMR results from across-channel comparisons of temporal envelopes. Alternatively, it has been suggested that analysis of the output of a broad initial predetection filter, which encompasses frequencies generally thought to fall into disparate auditory filters, can account for certain aspects of CMR (Berg, 1996). However, Buss et al. (1998) and Buss and Hall (1998) provided evidence against such a broad predetection filter; their results were, instead, consistent with an initial stage of auditory (bandpass) filtering. Other studies have proposed that within-channel cues, i.e., information from only the one auditory channel tuned to the signal frequency, can account for a considerable part of the effect in some conditions, which means that within-channel processing can lead to an overestimation of "true" across-channel CMR (e.g. Schooneveldt and Moore, 1987). This was supported by simulations of data from the band-widening experiment, using a modulation filterbank analysis of the stimuli at the output of the auditory filter tuned to the signal frequency (Verhey et al., 1999). Additionally, for the CMR

experiments using flanking bands, McFadden (1986) pointed out that it is imprecise to assume that one channel is receiving only the on-frequency band plus signal and another channel is receiving only the flanking band. Often, the two bands will be incompletely resolved. When this happens, the resulting waveform may contain envelope fluctuations resulting from beats between the carrier frequencies of the on-frequency and the flanker bands. These beats can facilitate signal detection without across-channel comparisons being involved. Thus, at least part of the masking release can be explained in terms of the use of within-channel rather than across-channel cues. Taken together, across-channel CMR appears to be a robust, but relatively small effect, which was found in monotic and dichotic conditions.

A recent study on effects of auditory grouping on CMR (see Chapter 4 and Dau et al., 2009) supported two forms of CMR. In their study, the effects of introducing a gating asynchrony between on-frequency and flanker bands or a stream of preceding (precursor) or following (postcursor) flanker bands were studied for conditions of CMR. Using widely (one octave) spaced flanking bands, CMR effects were eliminated by introducing a gating asynchrony and by introducing the pre- or postcursor flanking bands. Using narrowly spaced flanking bands (one-sixth octave), CMR was not affected by any of the stimulus manipulations. Their results supported the hypothesis that one form of CMR is based on within-channel mechanisms (Schooneveldt and Moore, 1987; Verhey et al., 1999), determined by the envelope statistics. The fact that this effect was not susceptible to manipulations by auditory grouping constraints is in line with the assumption that the mechanism is peripheral in nature, based on the physical interaction of stimulus components within an auditory channel. The other form of CMR, mainly based on true across-channel comparisons, appeared to be dependent on auditory grouping constraints, consistent with the results from Grose and Hall (1993).

Several hypotheses have been suggested about the nature of the across-channel mechanism underlying CMR. One hypothesis is based on the assumption that the addition of the signal to the on-frequency masker band leads to a change in the modulation depth in the auditory filter centered around the signal frequency. By comparing this modulation depth to that of other auditory filters for which the modulation depth is unaltered, subjects would increase their sensitivity to the presence

of the signal (Hall, 1986). A different explanation for CMR was proposed by Buus (1985), who suggested that the comodulated flanker band(s) provide valuable information about the moments at which the masker band has a relatively low energy. By attributing more weights to these valleys in the masker, the effective signal-to-noise ratio increases and detection improves. This mechanism was called "listening in the valleys". Also proposed by Buus (1985) was an equalization-cancellation (EC) mechanism, originally introduced by Durlach (1963), to account for various binaural masking release data. According to this mechanism, the envelope of the masker and flanking band are first equalized and then subtracted. The output of such a mechanism might have a considerable increase in the signal-to-noise ratio provided that the masker and the flanking bands are comodulated.

A fourth mechanism has been proposed by Richards (1987), where it was assumed that the cross *covariance* between the envelopes of the masker and the flanking bands is used for signal detection. The envelope cross-covariance decreases when adding a signal to the masker and this cue might be used by the human auditory system. However, this model was later rejected because it was not compatible with experimental data by Eddins and Wright (1994). They used two 100% sinusoidally amplitude modulated sinusoids of different frequencies, and the subjects had to detect the in-phase addition of a sinusoid to one of the SAM sinusoids. The cross-covariance is not changed even though the modulation pattern is altered by the addition of the sinusoid. Thus, if changes in the cross-covariance were essential for receiving CMR, this type of stimulus should not lead to a CMR. This, however, was in contrast to their data, which clearly showed CMR.

Later, van de Par (1998) and van de Par and Kohlrausch (1998) found that CMR can better be described in terms of an envelope cross *correlation* mechanism than an envelope cross-covariance mechanism. Their study was motivated by earlier findings by Bernstein and Trahiotis (1996) which showed that cross-correlation was more successful than cross-covariance when studying binaural detection phenomena. At high frequencies where these experiments were carried out, similar mechanisms may indeed underly monaural CMR and binaural masking level differences (BMLD, van de Par and Kohlrausch, 1998). Moreover, the EC mechanism which has been used to

account for BMLD, was shown to be equivalent to a decision mechanism based on cross-correlation (Domnitz and Colburn, 1977; Green, 1992).

While potential mechanisms underlying CMR have been discussed in several studies, simulations that quantify the (relative) contributions of across- versus within-channel processing in different types of experiments have not been provided. The purpose of the present study was therefore to develop and evaluate a model that accounts for both effects in CMR. The modulation-filterbank model by Dau et al. (1997a,b) was considered as the modeling framework. This model was used earlier to analyze within-channel cues in CMR obtained in the band-widening experiment (Verhey et al., 1999), and applied to a variety of other detection and masking conditions, including tone-in-noise detection, modulation detection, and forward masking. In the Verhey et al. (1999) study, the model was exclusively tested in the band-widening experiment of CMR. The results from the simulations, performed only in the auditory channel tuned to the signal frequency, suggested that essentially no across-channel processing is involved in this type of CMR condition. Instead, temporal within-channel cues such as beating between components, evaluated by the modulation filterbank model, appear to account for the masking release in the model simulations. However, since the model does not contain any explicit across-channel processing, it will not be able to account for any "true" across-channel CMR. In the present study, an EC-based circuit was integrated into an extended version of the modulation filterbank model whereby the EC processing was assumed to take place at the level of the internal representation of the stimuli *after* modulation filtering.

First, the structure of the across-channel modulation filterbank model is described. The model is then evaluated in several experimental conditions: (i) CMR with four widely spaced flanker bands to study pure across-channel CMR (Experiment 1), (ii) CMR with one flanking band varying in frequency in order to study the transition between conditions dominated by within-channel processing and those dominated by across-channel processing (Experiment 2), and (iii) CMR obtained in the band-widening paradigm in order to study the contribution of across-channel processing in a condition which always includes within-channel processing (Experiment 3). For direct comparison, experimental data were obtained in the same conditions with exactly the

same stimuli and using exactly the same threshold algorithm as in the simulations. The results and implications for further modeling work are discussed.

2.2 Model

The model presented here is based on the monaural detection model of Dau et al. (1997a). The original model was designed to account for signal detection data in various psychoacoustic conditions. It has proven successful in predicting data from spectral and spectro-temporal masking (Verhey et al., 1999; Derleth and Dau, 2000; Verhey, 2002), non-simultaneous masking (Dau et al., 1996, 1997a; Derleth et al., 2001), and modulation detection and masking (Dau et al., 1997a,b, 1999; Ewert and Dau, 2004). In the meantime, an additional model of amplitude modulation (AM) processing, the envelope power spectrum model (EPSM) has been developed (Ewert and Dau, 2000; Ewert et al., 2002), based on Viemeister (1979) and Dau et al. (1999). The EPSM has a much simpler structure than the above mentioned processing model. It is similar to Viemeister's (1979) leaky-integrator model but assumes modulation bandpass filters instead of a single modulation lowpass filter. It consists of only three stages: Hilbert-envelope extraction, modulation bandpass filtering, and a decision stage based on the long-term, mean integrated envelope power. This model does not include any effects of peripheral filtering and adaptation, and timing information (as reflected in the envelope phase and modulation beatings) is neglected. While the EPSM demonstrated in a straight-forward and intuitive way the need for modulation-frequency selective processing and can account for modulation masking data, it is conceptually less general than the perception model (Dau et al., 1996, 1997a).

The model as described in Dau et al. (1997a), which forms the basis for the model developed here, consists of the following steps: Peripheral filtering, envelope extraction, nonlinear adaptation, modulation filtering, and an optimal detector as the decision device. To simulate the bandpass characteristic of the basilar membrane, the gammatone filterbank (Patterson et al., 1987) is used. At the output of each peripheral filter, the model includes half-wave rectification and low-pass filtering at 1 kHz. While the fine structure is preserved for low frequencies, for high center frequencies this stage essentially preserves the envelope of the signal. Effects of adaptation are

simulated by a nonlinear adaptation circuit (Püschel, 1988; Dau et al., 1996). For a stationary input stimulus, this stage creates a compression close to logarithmic. With regard to the transformation of envelope fluctuations, the adaptation stage transforms the AM depth of input fluctuations with rates higher than about 2 Hz almost linearly. The stimuli at the output of the adaptation stage for each channel are then processed by a linear modulation filterbank. The lowest modulation filter is a second-order lowpass filter with a cutoff frequency of 2.5 Hz. For frequencies above 5 Hz there is an array of bandpass filters with a quality factor $Q = 2$. Modulation filters with a center frequency above 10 Hz only output the Hilbert envelope of the modulation filters, introducing a nonlinearity into the modulation processing through which the phase of the envelope is not preserved for the filters above 10 Hz. To model a limit of resolution, an internal noise with a constant variance is added to the output of each modulation filter. In the decision process, a stored, normalized temporal representation of the signal to be detected (the template) is compared with the actual activity pattern by calculating the cross-correlation between the two temporal patterns (Dau et al., 1996, 1997a). This is comparable to a "matched filtering" process (Green and Swets, 1966).

For the processing of arbitrary input stimuli, the function of the model can be considered as being separated in two (parallel) paths: (i) The stimulus representation after nonlinear adaptation is low-pass filtered at a cut-off frequency of 2.5 Hz, thereby essentially extracting the stimulus energy. With this processing alone, the model would be acting similarly to a power spectrum model (e.g., Patterson and Moore, 1986) and would account for certain aspects of spectral masking data (Derleth and Dau, 2000). (ii) The bank of modulation bandpass filters captures the dynamic properties of the stimulus. It is expected that, in the model, a hypothetical process underlying across-channel CMR would use the output of the bandpass modulation filters. So far, however, the model in its original form does not contain any explicit across-channel processing and therefore fails to produce "true" across-channel CMR.

The present study introduces an explicit across-channel mechanism into the model. Figure 2.1 illustrates the model used in the present study. The modification of the model is comparable to the EC mechanism of Durlach's model (Durlach, 1960, 1963) for describing binaural masking level differences (BMLDs). However, while the EC mechanism in the original (binaural) model is applied essentially to the stimulus

waveforms, and jitter is provided in the level and time domains in order to limit the resolution in the model, the (monaural) EC process in the current model is applied at a much later stage of auditory processing, and no additional limitations are introduced. In contrast to the original binaural EC model, it is assumed here that the limitations for performance are already included in the processing stages prior to the EC process.

The essential aspects of this approach are first illustrated for only two peripheral channels, i.e., using a channel centered at the on-frequency band including the signal, and a channel centered at one remote flanking band.

The across-channel processing within the model is assumed to occur at the output of all (bandpass) modulation channels tuned to frequencies at and above 5 Hz, which is the center frequency of the lowest modulation filter. The individual modulation filter outputs at the flanking band are subtracted from the corresponding outputs at the on-frequency channel. This process is denoted as cancellation in Figure 2.1. The outputs of the lowpass filters in the different peripheral channels remain unaffected. The low-pass filtered outputs as well as the difference representations after modulation bandpass filtering are subjected to the decision stage, the optimal detector, which assumes independent observations for the different inputs. The specific case with only one flanking band does not require an equalization stage.

Typically, more than one flanking band will be presented. The generalized mechanism for the multi-channel case is indicated in Fig. 2.2. Here, the weighted sum of the activity of the flanking bands is computed and subtracted from the on-frequency channel. Calculating the weighted sum can be considered as equalization process, since it equalizes the summed activity in the different flanking bands with regard to the on-frequency band. The subtraction refers to a cancellation process as in the case with only one flanking band (Fig. 2.1).

In Fig. 2.2, a situation with N flanking bands and one on-frequency band is assumed. Here, the EC mechanism acts on the N peripheral channels, denoted as PC1, PC2, ..., PCN. PCX indicates the channel centered at the on-frequency band. For simplicity, only the output of the j-th modulation filter $s_{jn}(t)$ in the different peripheral channels ($n = 1...N$) is indicated in the figure. The outputs of all other modulation filters are processed in the same way. The output $s_j(t)$ of the EC mechanism for N channels at the j-th modulation filter can be expressed as

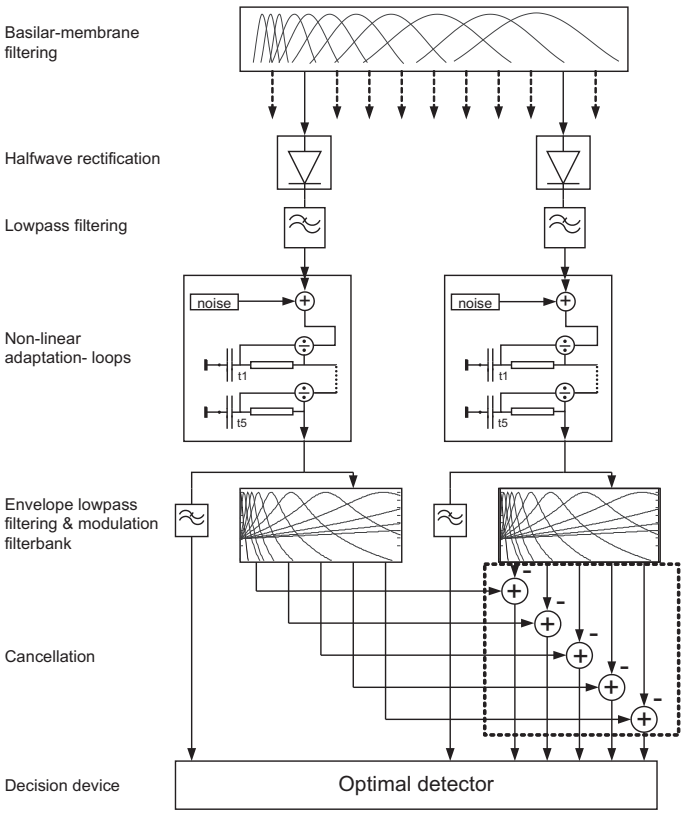

Figure 2.1: Block diagram of the across-channel modulation filterbank model. The signals are filtered by the gammatone filterbank, half-wave rectified and low-pass filtered at 1 kHz, and subjected to adaptation. The adapted signal is then filtered by a modulation bandpass filterbank and a separate lowpass filter (at 2.5 Hz) at the output of each auditory filter. At the output of the individual modulation bandpass filters, the activity at the flanking bands is averaged across the flankers (E-process) and subtracted from the corresponding activity at the on-frequency band (C-process), illustrated here with only one flanking band and highlighted in the dashed box. The output activity is added to internal noise and finally subjected to an optimal detector as decision device.

$$s_j(t) = s_{jx}(t) - c_j(t) = s_{jx}(t) - \frac{\sum_{\substack{i=1 \\ i \neq x}}^{N} w_i \, a_i \, s_{ji}}{\sum_{\substack{i=1 \\ i \neq x}}^{N} a_i} \,, \tag{2.1}$$

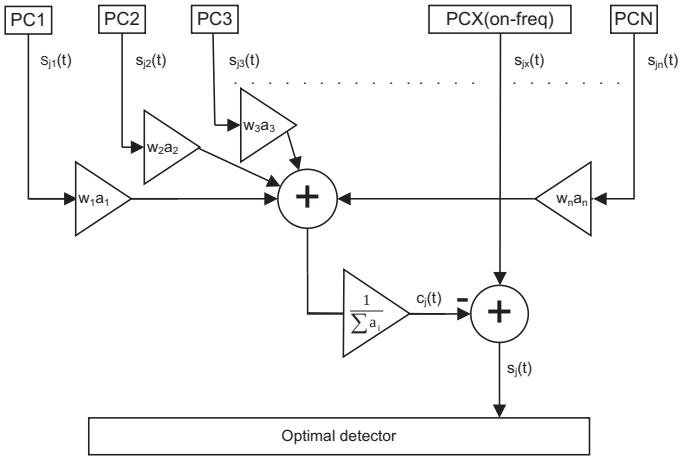

Figure 2.2: Simplified block diagram of the across-channel EC process in the perceptual model for N peripheral channels PC1,...,PCN. Only one modulation filter at each peripheral channel is shown.

where the index x denotes the peripheral channel (PCX) tuned to the on-frequency band and $c_j(t)$ represents the cancellation term. The contributions $s_{j1}, s_{j2}, ..., s_{jN}$ are weighted by the factors $a_1, a_2, ..., a_N$. The weights a_i equal the root-mean-square (rms) of the lowpass filter output in the channels PCi ($i = 1, .., N$). Since the rms value reflects the average energy of a signal, a_i equals the average energy in the i-th peripheral channel. Thus, the weighting with a_i means that the channels that are excited by more input stimulus energy are emphasized relative to the filters which are excited by less. Specifically, filters without excitation by the stimulus do not contribute at all to the cancellation term $c_j(t)$. The cancellation term includes a normalization by the factor $\sum_{\substack{i=1 \\ i \neq x}}^{N} a_i$ that is proportional to the overall energy of the stimuli in all peripheral channels except the on-frequency channel. In order to make sure that the EC stage operates *across* channels and does not subtract much signal information from the signal channel, the off-frequency weight w_i was introduced. In the current implementation, w_i was set to zero if the overlap between the magnitude transfer function of the auditory channels at PCi and PCX is above a certain limit, and was set to one otherwise. The overlap of the filter transfer functions was calculated during

the design phase of the model as the correlation value of broadband noise at the output of the two respective filters. The limit was chosen to be a correlation value of 5%. In this way, auditory filters tuned at and very close to the signal frequency were not considered in the EC process. The weights w_i ensure that, for example, in the case of a broadband noise as input, the stimuli in the channels contributing to the cancellation term are statistically independent from the excitatory on-frequency channel. Thus the EC mechanism in the model can be regarded as a "true" across-channel process.

In the most general version of the model, the EC process would be considered in all peripheral channels covering the whole audible frequency range, with each of the channels being regarded as a potential signal channel and with all respective surrounding channels being included in the cancellation term. In the simulations of the present study, however, the model was "told" in advance which was the signal frequency and thus which was the on-frequency channel. All remaining channels in the range from 500 to 6000 Hz were considered as the cancellation channels. This simplification is based on the assumption that the best signal-to-noise ratio is expected to be in the channel tuned to the signal and that detection is mainly based on this single channel (including the information from the other channels contained in the cancellation term of the EC process). An additional simplification was made in conditions when the stimulus was sparsely represented along the peripheral channels as, e.g., in the case of widely spaced narrow-band flankers in experiment 1. In this case, only channels tuned to the frequencies of the flanker bands were considered. The off-frequency weights w_i were then equal to one. If all flanker bands have equal energy (as in experiment 1), all a_i have the same value a. The cancellation term $c_j(t)$ in Eqn. 2.1 can then be simplified to:

$$c_j(t) = \frac{\sum_{\substack{i=1 \\ i \neq x}}^{N} a\, s_{ji}}{\sum_{i=1 i \neq x}^{N} a} = \frac{\sum_{i=1 i \neq x}^{N} s_{ji}}{N-1} \tag{2.2}$$

and becomes the average over the number of flanking bands.

2.3 Method

2.3.1 Subjects

Four normal-hearing subjects participated in each experiment. Their ages ranged from 23 to 41 years. All subjects had experience in other psychoacoustic experiments. The authors TP and TD participated in the experiment. The other two subjects were paid for their participation on an hourly basis.

2.3.2 Apparatus and stimuli

The subjects were seated in a double-walled, sound attenuating booth and listened via Sennheiser HD580 headphones. Signal generation and presentation during the experiments were computer controlled using the AFC software package for MATLAB, developed at Universität Oldenburg and DTU. All stimuli were generated digitally on an IBM compatible PC and were then converted to analog signals by a high-quality 32-bit soundcard (RME DIGI-96PAD) at a sampling rate of 32 kHz. Three CMR experiments were performed where the subject's task was to detect a tone in the presence of one or more noise masker bands. The specific stimuli will be described in the respective experiments (Experiments 1-3).

2.3.3 Procedure

A three-interval, three-alternative forced-choice paradigm was used to measure detection thresholds. A two-down, one-up procedure was used to estimate the 70.7% correct point of the psychometric function (Levitt, 1971). Subjects had to identify the one randomly chosen interval containing the signal. Subjects received visual feedback if the response was correct. The three observation intervals were separated by 500 ms of silence. The initial step size for the signal level was 4 dB and after every second reversal of the level adjustment the step size was halved until the step size of 1 dB was reached. The mean of the signal level at the last six reversal was calculated and regarded as the masked threshold value. For each stimulus configuration and subject, four masked threshold values were measured. The mean of these values was calculated

and taken as the final threshold. For the model simulations the identical procedure and the same AFC framework as in the experiments were used.

2.4 Experiment 1: CMR with four flanking bands

2.4.1 Rationale

The first experiment was designed to investigate "true" across-channel CMR, where within-channel processing does not contribute. Four flanking bands with a spectral separation of one octave were used such that within-channel contributions to CMR can be assumed to be negligible at the (medium) sound pressure levels used in this experiment.

2.4.2 Stimuli

The signal was a 1000-Hz pure tone. The masker consisted of five bands of noise which were centered at 250, 500, 1000, 2000 and 4000 Hz, thus covering a frequency range of 4 octaves. Signal and masker had the same duration of 187.5 ms. 20-ms raised-cosine ramps were applied to the stimuli. Signal threshold was measured as a function of the bandwidth of the masker, which was 25, 50, 100 or 200 Hz. The masker bands were generated in the time domain, transformed to the frequency domain by Fourier transform where they were restricted to the desired bandwidth, and finally transformed back to the time domain by inverse Fourier transform. In the reference condition, the envelopes of the five bands were uncorrelated with each other. In the comodulated condition, the on-frequency noise masker was shifted to the center frequencies of the flanking bands in the Fourier domain, such that the envelopes of the different bands were fully correlated with each other. The presentation level of each of the maskers was 60 dB SPL.

2.4.3 Results

Figure 2.3 shows the results of the experiment. Masked thresholds are plotted as a function of the masker bandwidth. The open symbols represent the experimental

data, averaged across subjects. The circles and squares show the results for the uncorrelated and comodulated conditions, respectively. The right panel of Fig. 2.3 shows the amount of CMR, i.e., the difference between the uncorrelated and comodulated thresholds. There is a significant CMR effect of 4-5 dB for the small noise bandwidths of 25 and 50 Hz (one-way ANOVA: $F(1,18) = 38.59, p < 0.001$ and $F(1,18) = 32.18, p < 0.001$), while no significant CMR was found for the larger bandwidths of 100 and 200 Hz (one-way ANOVA: $F(1,18) = 1.67, p = 0.21$ and $F(1,18) = 0.02, p = 0.89$) where statistical significance here and in the following is defined as having $p < 0.01$. Thus, even though four flanking bands were used, the obtained CMR is relatively small compared to the results typically found with narrow spacing between the signal and flanking bands (see experiment 2) or in the band-widening CMR paradigm (see experiment 3). The results are consistent with results from previous studies (e.g., Moore and Emmerich, 1990), showing that CMR is restricted to narrowband noises with bandwidth smaller than 50 Hz. This indicates that across-channel CMR is a phenomenon that occurs only when the masker is dominated by relatively slow envelope fluctuations. The modulation spectrum of bandpass noise is directly related to the bandwidth of the noise (e.g., Lawson and Uhlenbeck, 1950; Dau et al., 1997a). The rate of modulations will range up to the bandwidth of the noise, Δf.

The filled symbols in Fig. 2.3 show the simulations obtained with the processing model described in Sect. II. The simulations represent average thresholds of 10 repetitions for each experimental condition. The model predicts slightly elevated overall thresholds (2-3 dB) and larger standard deviations in comparison to the empirical data. For the bandwidths 25 and 50 Hz, the model predicts a significant mean CMR effect of about 4 dB (one-way ANOVA: $F(1) = 15.38, p < 0.001$ and $F(1,18) = 16.91, p < 0.001$, respectively). It does not produces a significant amount of CMR for the 100 and 200-Hz bandwidths (one-way ANOVA: $F(1,18) = 6.48, p = 0.02$ and $F(1,18) = 6.29, p = 0.02$).

2.4.4 Model analysis

The following describes how the EC-mechanism affects the signal processing of the stimuli in the model. Since the EC-process typically leads to a lower threshold

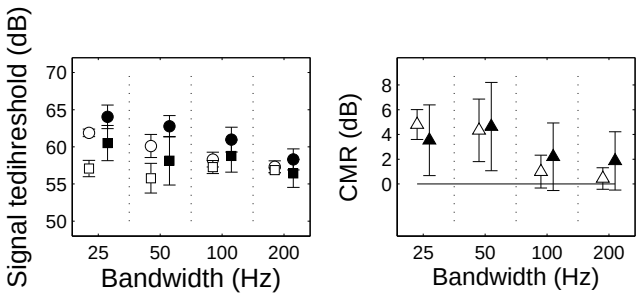

Figure 2.3: Left panel: Detection thresholds for the 1-kHz tone in the presence of five noise bands as a function of the bandwidth of the noises. Open symbols indicate average experimental data and filled symbols show simulation results. Circles and squares represent results for the uncorrelated and comodulated conditions, respectively. Right panel: CMR effect for the conditions of the left panel.

in the comodulated condition compared to the uncorrelated condition, this should be reflected in the model's internal representations of the stimuli. As an example, the upper left panel of Fig. 2.4 shows the internal representation of a single 25-Hz wide (comodulated) noise masker centered at 1 kHz. The outputs of the modulation filters are shown separately in the subpanels, including the modulation lowpass filter (indicated as 0 Hz), and the bandpass filters tuned to 5, 10, 17, 28, 46, 77, 129, and 214 Hz. The solid curves show the output obtained without EC-process, i.e., when using the original model's (Dau et al., 1997a) preprocessing. The dashed curves show the output when the EC process was included, i.e., after subtracting the average activity of the four flanking bands from the on-frequency band. As expected, the output representation (for the comodulated noise bands) after the EC process is reduced in amplitude compared to the result without the EC process. Note that modulation channels tuned to frequencies higher than the bandwidth of the noise (25 Hz) are activated as well, mainly reflecting the response to the onset of the adapted envelope of the stimulus.

As described in Sect. II and in previous publications (Dau et al., 1996, 1997a), in the simulations, the internal representation of the noise is subtracted from the internal representation (either noise alone or signal plus noise) of each of the three intervals and then cross-correlated with the template. The template represents the normalized

difference between the internal representation of the noise plus *supra*-threshold signal and the noise-alone representation. The upper right panel of Fig. 2.4 shows the model's template using the same 25-Hz wide noise (as used for the illustration of the reference) to which a supra-threshold 1-kHz tone was added. As for the reference representation, the individual modulation filter outputs are indicated in the subpanels. In the case of the template, there is essentially no difference between the situation with and without EC-process since the internal representation of the template is dominated by the presence of the signal.

In order to evaluate the function of the EC-mechanism, the two lower panels of Fig. 2.4 show a statistical analysis of the cross-correlation between noise-alone representation and template (triangles), and between noise-plus-actual-signal representation and template (circles) including the EC-mechanism in the processing. The histograms of the cross-correlation coefficient are shown for the output of the same individual modulation filters as considered in the top panels. The "actual" signal level was chosen to be the simulated signal level at threshold (from Fig. 3, random condition). For the template, the same supra-threshold level (85 dB) was used as in the simulations. The lower left panel shows the results for the random noise condition at the output of the EC process. Since the signal level was chosen to be at detection threshold, the distributions are just separable (in terms of signal detection theory). The right panel shows the corresponding results for the comodulated condition. Here, the EC mechanism causes a strong sharpening of the distribution of correlations in the reference interval while the distributions in the signal interval remain essentially unaffected. This corresponds to an increased sensitivity and a decreased detection threshold in the simulations in the comodulated condition relative to the random condition, and represents the "noise reduction" caused by the EC mechanism. Without the EC-mechanism, the histograms would be similar in the random and comodulated condition.

The comparison of the histograms at the output of the different modulation filters suggests that all modulation filters contribute to signal detection (also those tuned to modulation frequencies higher than the noise bandwidth of 25 Hz). In other words, the decision in the model does not seem to be based on the activity at the output of only one or a few particular modulation filters. This is different from the

situation in conditions of within-channel CMR, at least in the framework of the current model, where modulation cues like beatings between on-frequency and flanker bands components become effective and activate specific modulation filters in the signal interval (see the corresponding analysis in experiment 2 further below). In the EC model, a supra-threshold signal does not produce any specific modulation pattern that could be used as cue. The EC mechanism therefore does not lead to an enhancement of specific cues which would be reflected by different templates for the same condition with or without EC mechanism. The EC mechanism rather suppresses the noise fluctuations in the modulation filters, thereby enhancing signal detection.

Since the outputs of all bandpass modulation filters contribute to the function of the EC mechanism in the model, the question remains whether a modulation filterbank is necessary for the occurrence of CMR. To address this question, additional simulations were carried out with alternative modulation filtering stages: (i) A process referred to as "DC/AC" which separates the DC-component of the Hilbert envelope spectrum from the remaining (AC) spectrum, (ii) a combination of a second-order Butterworth low-pass and a high-pass filter with cutoff-frequencies of 2.5 Hz, referred to as "LH", (iii) a combination of the same low-pass filter at 2.5 Hz combined with a single bandpass filter centered at 5 Hz with a bandwidth of 5 Hz, referred to as process "LB5", and (iv) the same as (iii) but with a bandpass filter tuned to 50 Hz and a bandwidth of 25 Hz (Q=2; referred to as "LB50"). The EC-process was applied to the AC-coupled output in DC/AC, the output of the high-pass filter in LH, and the output of the single bandpass filters in LB5 and LB50, respectively. Figure 2.5 (left panel) shows the corresponding simulations obtained with the different processing schemes for the random and the comodulated noise conditions using the same symbols as in Fig. 2.3. The right panel shows the amount of CMR for the different schemes. The result obtained with the complete modulation filterbank, referred to as "MF", was replotted from Fig. 2.3 for direct comparison.

The DC/AC and LH processes do not produce any CMR (one-way ANOVA: $F(1,18) = 1.51, p = 0.24$ for DC/AC, $F(1,18) = 0.16, p = 0.68$ for LH). In contrast, the processing of LB5 and LB50 produce a significant CMR effect of about 4 dB (one-way ANOVA: $F(1,18) = 30.96, p < 0.001$ and $F(1,18) = 15.4, p < 0.001$) which corresponds to the simulation obtained with the complete

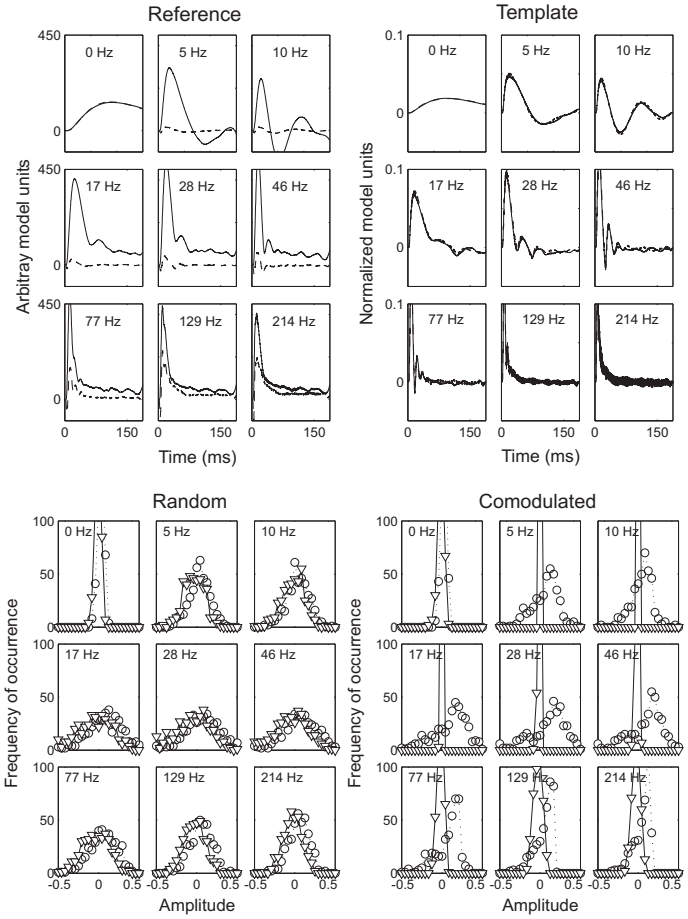

Figure 2.4: Simulated internal representations at the output of the modulation filters (indicated by the center frequencies in the sub-panels) in the on-frequency (peripheral) channel. Solid curves show outputs without EC process, dashed curves show results after the EC process. Left upper panel: Internal representation of (modulated) noise alone (i.e. no signal was added). Right upper panel: Internal representation of the template, i.e. the normalized difference between noise plus supra-threshold signal representation and noise alone representation. The lower panels show histograms of the cross-correlation coefficients between the noise-alone representation and template (triangles, solid line), and between the noise-plus-actual-signal representation and template (circles, dotted line), for the same individual modulation filters as considered in the top panels. This is shown for the random condition (left) and the comodulated condition (right), with EC mechanism applied.

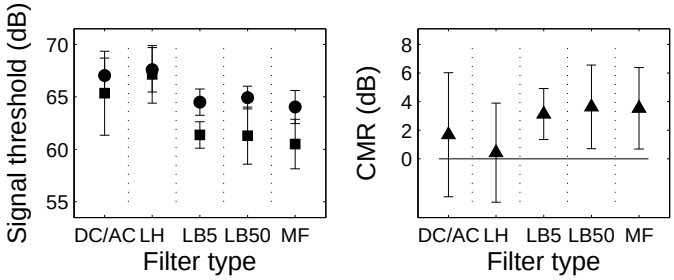

Figure 2.5: Left: Signal thresholds obtained with the filter types DC/AC, LH, LB5, LB50, as defined in the main text and the complete modulation filterbank (MF). Circles and squares show results for random and comodulated noise, respectively. Right: Amount of CMR for the different filter types.

modulation filterbank MF (one-way ANOVA: $F(1,18) = 38.59, p < 0.001$). Thus, within the model, across-channel CMR can only be produced if the stimulus after peripheral filtering, envelope extraction and adaptation is actually processed by frequency-selective (modulation) filters, whereby each individual filter would already be sufficient to produce significant CMR. The effect, however, disappears if only one broad (5-150 Hz) modulation bandpass filter is considered (not shown explicitly). The reason for the behavior in the model is that the input to the modulation filtering process, the adapted envelope, shows an onset response. This onset produces an excitation also at higher modulation frequencies. The EC process is only effective if the output of the modulation filtering process leads to a reasonable correlation between the flanking band and the signal band representations. This is only the case after (modulation) bandpass filtering, and cannot be obtained for the "broadband" schemes DC/AC and LH considered above. It is not clear, of course, to what extent the mechanisms in the real system are related to the ones proposed here on the basis of the model. The intention of the above analysis was to elucidate the functioning of the EC-process of the proposed model.

In summary, the data from Experiment 1 confirm results from previous studies that across-channel processing in CMR is robust but small (even when several flanking bands are involved). Across-channel CMR is only observable at small bandwidths (below about 50 Hz), i.e., when the envelope fluctuations inherent in the stimuli are

relatively slow. The simulations indicate that across-channel CMR can be accounted for quantitatively if an EC-like mechanism is introduced at the output of a modulation frequency-selective process.

2.5 Experiment 2: CMR with one flanking band varying in frequency

2.5.1 Rationale

This experiment investigates the transition between conditions where exclusively across-channel mechanisms determine CMR and those where primarily within-channel mechanisms generate CMR. Only one flanking band was used here, as in the study by Schooneveldt and Moore (1987). The amount of CMR was measured and simulated as a function of the spectral separation between the flanking and the on-frequency band. While for large separations of one octave or greater, CMR cannot be expected to exceed 2-4 dB, masking releases of about 14 dB and higher were observed in previous studies for separations of less than $1/10$ octave where within-channel processing provides the most effective detection cues (Schooneveldt and Moore, 1987). A successful model of CMR needs to account for both within- and across-channel components.

2.5.2 Stimuli

The stimuli were similar to some of those used in Schooneveldt and Moore (1987). The signal was a 2000-Hz tone. The on-frequency masker was a 25-Hz wide band of noise centered at the signal frequency. The flanking band had the same bandwidth as the on-frequency band and was centered at 1000, 1400, 1800, 1900, 2100, 2200, 2600 or 3000 Hz, corresponding to frequency ratios between flanking band and on-frequency band of 0.5, 0.7, 0.9, 0.95, 1.05, 1.1, 1.3, and 1.5. In contrast to the study by Schooneveldt and Moore (1987), the flanking band was not presented directly at the signal frequency or very close to it. The two noise bands were either uncorrelated or comodulated. As in Schooneveldt and Moore (1987) each band was produced by

multiplying a sinusoid at the center frequency with a low-pass noise with a cutoff frequency of 12.5 Hz. In the comodulated condition, the noise bands were produced by multiplying the different sinusoids with an identical low-pass noise whereby a new noise was generated for each interval. Each band had an overall level of 67 dB SPL.

2.5.3 Results and model analysis

Panel (a) of Fig. 2.6 shows average data for the uncorrelated (open circles) and the comodulated (open squares) conditions. The signal threshold is plotted as a function of the ratio between flanking-band and signal frequency. The difference in threshold between uncorrelated and comodulated conditions, i.e., the amount of CMR, reaches 12-14 dB when flanker and signal frequency are close to each other (with ratios between 0.9 and 1.1). For large separations between on-frequency and flanking band, the data show a slight asymmetry: CMR of 3-4 dB in the presence of the high-frequency flankers and 5-6 dB for flanking bands presented at low frequencies. The data agree well with the results of Schooneveldt and Moore (1987).

Panel (b) of Fig. 2.6 shows the simulations obtained with the present model. As described in Sect. II, the EC mechanism was applied in all filters that overlap less than 5% with the on-frequency gammatone filter, i.e., in all channels except the two closest ones on both sides of the on-frequency channel. In this particular experiment, this means that the flanker bands were maximally contributing to the cancellation term of the EC process at frequency ratios of 0.5, 0.7, 1.3, and 1.5. The model accounts for the relatively flat threshold function obtained in the uncorrelated condition. For flanking-band frequencies close to the signal frequency (at the frequency ratios 0.95 and 1.05), the model predicts a large amount of CMR that corresponds to that found in the experimental data. This component depends on beating of the carrier frequencies of the on-frequency and flanking bands. In the model this can be accounted for by the processing within the (peripheral) channel tuned to the signal frequency. The model detects changes in the envelope statistic due to the addition of the signal to the on-frequency band (see Verhey et al., 1999). This is effective for the comodulated condition while it does not provide additional detection cues in the uncorrelated condition. At very low and very high flanking band frequencies, the model predicts an average amount of CMR of about 3 dB which agrees well with the data at the

high flanker frequencies but is slightly less than the measured effect at the low flanker frequencies. The simulated 3-dB effect is the result of the EC mechanism in the model as can be seen from direct comparison with the results obtained without EC-circuit, shown in panel (c) of Fig. 2.6. As expected, without across-channel processing, no CMR is simulated at the large frequency separations between the on-frequency and the flanker band.

While certain aspects of the data can be described satisfactorily by the model, some other aspects can not. First, the simulated threshold function for the comodulated condition increases too steeply with increasing spectral distance from the signal. Second, the simulated amount of CMR for the lowest flanker frequencies is smaller than in the data. The reason for these discrepancies might be related to the shape of the magnitude transfer function of the peripheral filters used in the simulations. The gammatone filters are symmetrical on a linear frequency scale. However, it has been demonstrated that below its center frequency, the skirt of the human auditory filter broadens substantially with increasing stimulus level, and above its center frequency the skirt sharpens slightly with increasing level (Lutfi and Patterson, 1984; Moore and Glasberg, 1987). In order to illustrate effects of frequency selectivity on CMR in the framework of the current model, additional simulations were carried out using gammachirp filters (Irino and Patterson, 1997). The gammachirp filter has an asymmetric magnitude transfer function, and the degree of asymmetry in this filter is associated with stimulus level. The gammachirp filter was shown to provide a very good fit to human notched-noise masking data. Its impulse response is well defined and includes only one parameter more than the gammatone filter (see Eq. 2 in Irino and Patterson, 1997). In the present study, the impulse responses of the gammachirp filters were calculated for a level of 67 dB SPL. Here, as a simplification, the simulations were run with selected gammachirp filters tuned to the on-frequency band and the flanking band, respectively. A complete gammachirp filterbank with well defined level-dependent overlap has not been developed yet. As in the previous simulations with gammatone filters, the EC process was applied when the overlap between the off-frequency channel and the signal channel was below 5% which was only the case for the two outer data points (frequency ratios 0.5 and 1.5).

All other model parameters were kept the same as in the simulations with gammatone filters. The results are shown in panel (d) of Fig. 2.6.

The simulations with gammachirp filters account for many aspects of the experimental data. Due to the broader bandwidth of the gammachirp filter compared to the gammatone filter, within-channel cues become effective for a larger range of flanking-band frequencies. The plateau of low thresholds corresponds to that found in the data. At low and at high flanking-band frequencies, CMR amounts to 3-4 dB due to the EC processing in the model. However, the introduction of the gammachirp filter does not account for the slight asymmetry observed in the measured data, even though the transfer functions of the individual filters have an asymmetric shape. The simulated pattern for the comodulated condition actually produces the same thresholds at both ends. Still, the overall correspondence with the data is high. For direct comparison, panel (e) of the same figure shows the corresponding simulations without EC process. All data points except for the two outer ones are replotted from panel (d), since no EC process was applied for the inner data points in panel (d). As for the simulations with gammatone filters without the EC process, no CMR was obtained at the largest spectral separations between flanking and on-frequency band.

In order to illustrate the importance of within-channel cues available in the conditions where on-frequency band and flanking band are close to each other, Fig. 2.7 shows a statistical analysis similar to that presented in Experiment 1. Histograms of the cross-correlation between noise-alone representation and template (triangles) and noise-plus-actual-signal representation and template (circles) are shown for the outputs of the individual modulation filters. A frequency separation between on-frequency band and flanking band of 50 Hz was used for illustration. It can be seen in Fig. 2.7 that signal detection is mainly based on information at the output of the modulation filter tuned to about 46 Hz. Here, the mean of the signal distribution is clearly larger than that of the noise distribution. Thus, the addition of the signal to the masker causes changes in the internal representation of the stimuli such that it can effectively be evaluated in one (or only a few) modulation filters in this given task. This detection cue is qualitatively different from that discussed in connection with the across-channel process where signal detection was mainly based on the sharpening of the noise distribution at the output of the EC process in all modulation filters.

The results of Experiment 2 thus support the hypothesis that CMR has (at least) two components. One is restricted to flanking band frequencies around the signal frequency. This component reflects the use of within-channel cues (beating), rather than across-channel cues. The other component does not depend strongly on flanking-band frequency, but rather on across-channel cues. This across-channel component of CMR amounts to about 3 dB. While this has been proposed in earlier studies (e.g., Schooneveldt and Moore, 1987), the present study tried to provide quantitative modeling to test explicitly the (relative) contributions of within- and across-channel processing.

2.6 Experiment 3: CMR as a function of the masker bandwidth

2.6.1 Rationale

The third experiment considered the "classical" band-widening experiment where the masker was centered at the signal frequency and signal threshold was obtained as a function of the bandwidth of the masker. In contrast to the two previous experiments, the band-widening experiment does not allow for a separation between within- and across-channel processes; within-channel contributions will always contribute to CMR, even for large masker bandwidths when many auditory filters are excited by the noise. Verhey et al. (1999) showed that a single-channel analysis, which uses only the information in one peripheral channel tuned to the signal frequency, quantitatively accounts for the main CMR effect in the band-widening experiment. This suggested that across-channel processes are not involved or not effective in this class of CMR experiments, even though several auditory filters are excited by the noise. This was directly investigated here with the extended model that includes an explicit across-channel process while it keeps the ability to process within-channel cues, as shown in Experiment 2.

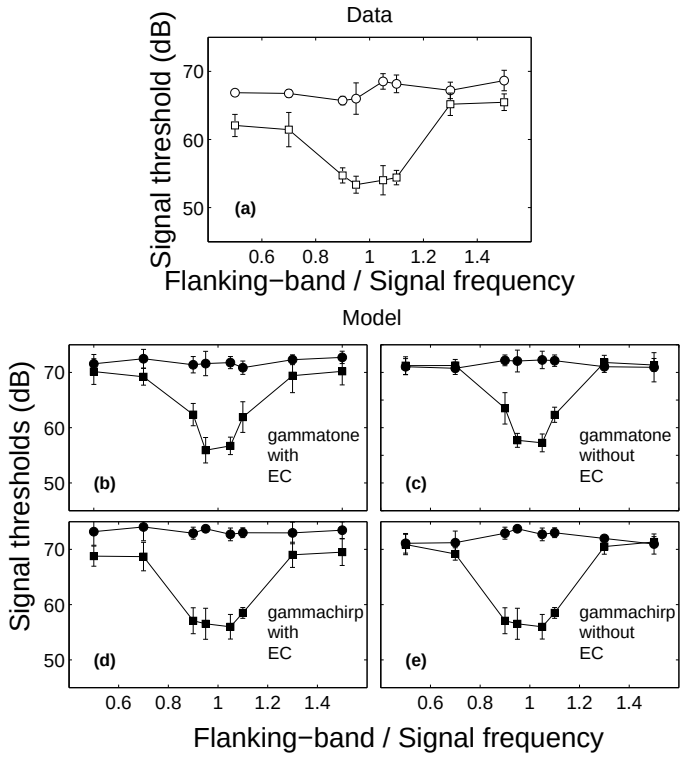

Figure 2.6: (a) Measured data averaged across subjects. Signal threshold for a 2-kHz tone in 25-Hz wide noise as a function of the spectral separation between on-frequency band and flanking band. Circles and squares show results for random and comodulated noise, respectively. (b) Simulations with the EC model shown in Fig. 2.1, using gammatone filters as the peripheral filtering stage. (c) Simulations with the same model, but with EC-process switched off. (d) Simulations as in (b) but with gammachirp filters. (e) Simulations with gammachirp filters, but with EC process switched off.

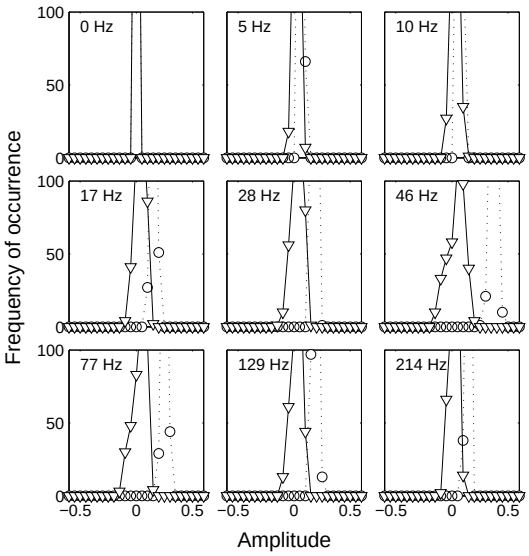

Figure 2.7: Histograms of the cross-correlation coefficients at the output of nine modulation filters in a condition of experiment 2 with 50 Hz separation between the on-frequency and flanking bands. Correlations for the reference (triangles, solid line) and reference-plus-signal (circles, dashed line) are shown for comodulated noise bands. The center frequency of the modulation filter is indicated within each panel. For the output of the modulation filter close to 50 Hz, the mean of the distributions is most different and the distributions are most separable in terms of signal detection.

2.6.2 Stimuli

The signal was a 300 ms long 2000-Hz tone. The masker was a band-limited noise centered at the signal frequency. The masker bandwidth was 50, 100, 200, 400, 1000 or 2000 Hz. The duration of the masker was 600 ms with 10-ms raised-cosine onset and offset ramps. The signal was temporally centered in the masker. Two types of maskers were used, as in the original experiments in Hall et al. (1984). One was a random noise with irregular and independent envelope fluctuations in different frequency regions. The comodulated noise was a random broadband noise which was modulated in amplitude at an irregular, low rate, and then restricted to the desired bandwidth. A low-pass noise with a cutoff at 50 Hz was used as a modulator.

Other studies have shown that for modulator bandwidths larger than 50 Hz, CMR decreases with increasing modulator bandwidth whereas it remains roughly constant for modulator bandwidth below this value (Schooneveldt and Moore, 1987; Carlyon and Stubbs, 1989). The modulation resulted in fluctuations in the amplitude of the noise which were the same in different frequency regions. The spectrum level of the bandpass noise was 30 dB, corresponding to overall levels of 47-63 dB SPL for the 50-2000 Hz bandwidth range.

2.6.3 Results

Figure 2.8 shows the results of the band-widening experiment. The left panel shows the experimental data, averaged across subjects. The signal threshold is plotted as a function of the masker bandwidth, for random noise (open circles) and comodulated noise (open squares). Consistent with the results from the earlier studies, for the random noise, the masked threshold first increases as the masker bandwidth is increased. Beyond a certain bandwidth (200 Hz in this case), the threshold no longer increases, but remains roughly constant. The increase of the threshold is caused by the fact that, up to the critical bandwidth, more noise passes through the auditory filter centered at the signal frequency, while beyond the critical bandwidth, the added noise falls outside the passband of the auditory filter. In contrast, for the comodulated noise, the threshold first stays constant and then decreases as the bandwidth is increased beyond about 200 Hz. The amount of CMR, defined as the difference in threshold between the random and comodulated conditions, is 12 dB for the largest bandwidth (2000 Hz).

The right panel of Fig. 2.8 shows the corresponding model simulations. For direct comparison, simulations are shown with EC process (dashed line) and without EC process (solid line). The two model versions essentially produce the same results. Thus, the across-channel processing does not generate any change in the overall amount of CMR in the framework of this model, not even at the largest masker bandwidths where several auditory channels are excited. Figure 2.9 shows the statistical analysis of the decision variable in the simulations, as in the first two experiments.

The comodulated condition with the broadest noise bandwidth (2000 Hz) was

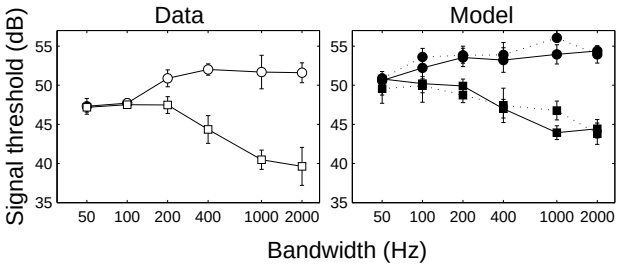

Figure 2.8: Left panel: Average signal thresholds for four subjects are plotted as a function of the masker bandwidth in random noise (circles) and comodulated noise (squares). Right panel: Simulated signal threshold of the model when the EC-mechanism is applied (dotted line) and when it is not applied (solid line). The modulator bandwidth was 50 Hz and the signal frequency was 2000 Hz.

considered for illustration with and without EC mechanism. At this bandwidth, the observed amount of CMR is maximal (12 dB). The analysis was carried out at a signal level of 55 dB which is about 10 dB above the simulated threshold in the comodulated condition. The left panel shows the distribution of the cross-correlation between noise-alone representation and template (triangles) and for the signal-plus-noise representation and template (circles) at the output of the single (peripheral) channel tuned to the signal frequency (single-channel analysis). It can be seen that there is a separation between the two distributions at several modulation filter outputs.

Since the bandwidth of the noise (also after peripheral filtering) is larger in this experimental condition than in the previous experiments, the variability of the envelope amplitude fluctuations is smaller, leading to the relatively sharp distributions. The right panel shows the analysis including the across-channel process in the model, i.e. a multi-channel simulation was carried out in this case where the cancellation term in the EC process was derived from the off-frequency channels. The envelope correlation across the different peripheral channels is *not sufficient* to effectively increase the signal-to-noise ratio at the output of the EC-process in the model. The EC process therefore does not contribute to signal detection in this type of experiment in the framework of the model.

These results therefore support the hypothesis that CMR obtained in the band-

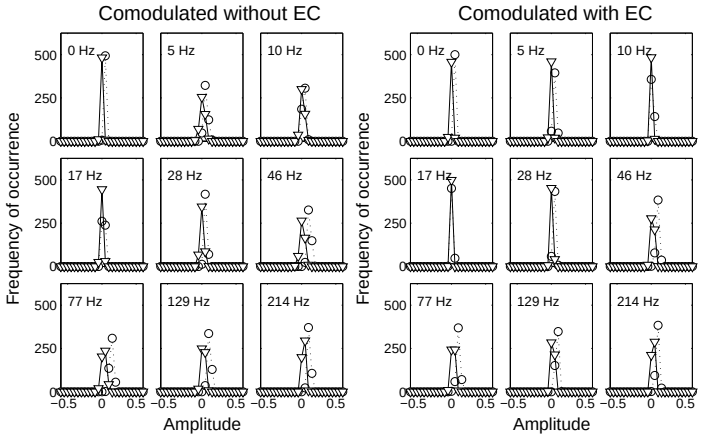

Figure 2.9: Histograms of the cross-correlation coefficients at the output of nine modulation filters for the comodulated conditions of Experiment 3 with a noise bandwidth of 2000 Hz. Left panel: Reference alone (triangles, solid line) and reference-plus-signal (circles, dashed line) for comodulated noises without EC. Right panel: Reference alone (triangles, solid line) and reference-plus-signal (circles, dashed line) for comodulated noises with EC process included.

widening paradigm is strongly dominated by within-channel processing and is not a result of across-channel processing.

2.7 Discussion

2.7.1 Within- versus across-channel processing

The modeling results of this study support the hypothesis that (at least) two mechanisms are contributing to what has been defined as CMR. The present model allows a distinction to be made between these two contributions. The simulations strongly support that one of the processes is based on within-channel mechanisms. Signal detection is based on the changes of the internal representation of the stimuli at the output of individual auditory filters – without the need for explicit across-frequency processing. The addition of the signal to the comodulated masker typically changes the (envelope) statistics of the stimuli significantly, while the changes are

much smaller or absent in the case of random noise maskers (Schooneveldt and Moore, 1987; Verhey et al., 1999). CMR resulting from within-channel contributions can be up to about 15 dB depending on the specific condition, and modulations (for example resulting from beatings between signal and masker components) up to several hundred Hz can serve as a cue for signal detection. A prerequisite for accounting for the full range of within-channel contributions to CMR is therefore a high sensitivity of the model to amplitude modulations (see also Verhey et al., 1999), as is the case for the modulation filterbank used in the present framework. Specifically, the modeling results suggest that a few individual modulation filters (at the output of the single peripheral channel at the signal frequency) can process the changes in the internal representation of the stimuli effectively.

The other form of CMR is based on "true" across-channel processing. This effect is also robust but relatively small (2-4 dB) and becomes only effective when narrowband noises with bandwidths below about 50 Hz are presented, i.e. when the envelope fluctuations of the noises vary relatively slowly. The EC model described in the present study makes specific assumptions about how envelope information at the output of different auditory channels might be processed. The EC process was assumed to take place at the output of each modulation bandpass filter. The effect of the EC process is that the variance of the external noise (originating from the masker) at the level of the internal representations after the EC process is reduced in the case of the comodulated noise condition. This leads to improved signal detection compared to the random noise condition. In the framework of the model, the detection cue is thus qualitatively very different from the situation where within-channel processing determines CMR.

It is clear that effects of nonlinear peripheral processing, such as the level-dependent auditory filter bandwidth, have an influence on the relative contributions of within- and across-channel processing to CMR. In fact, some of the effects that were considered as across-channel contributions in the past might become within-channel contribution with proper modeling of non-linear filters. For example, at very high stimulus levels where the auditory filter bandwidth is markedly increased (compared to the gammatone filters used in the present study), it can be expected that even in conditions with very broad spacing between the on-frequency band and the

flanking band(s), CMR might be dominated by within-channel contributions. Ernst and Verhey (2005) have shown that CMR over ranges of three octaves can be modeled as a suppression effect in a non-linear single-channel model, using the dual resonance nonlinear filter (DRNL) model (Meddis et al., 2001). In some of the conditions in their study, however, the level of the off-frequency flanker was much higher (up to 60 dB) than that of the on-frequency band. Although their results are not directly comparable to the experimental conditions used in the present study, it can be assumed that with proper modeling of the non-linear auditory filters even more signal configurations that have been considered as across-channel in the past might reveal a within-channel contribution. Our current definition of when across-channel processing is applied to a particular filter is based on the amount of overlap of its transfer function with that of the signal channel. This definition might be general enough to also apply to filters of different or varying shapes and to non-linear filters; the approach was successful when analyzing the results of Experiment 2, where individual gammachirp filters were considered. This, however, needs to be further investigated using a complete filterbank of filters with different shape or a nonlinear filterbank such as, e.g, a bank of DRNL filters or a gammachirp filterbank.

The observation that two conceptually different mechanisms define CMR is compatible with the results from studies on effects of auditory grouping on CMR (see e.g. Chapter 4 and Grose and Hall, 1993). When widely spaced flanking bands were used (as in the first experiment of the present study), CMR effects could be eliminated completely by introducing a gating asynchrony between the on-frequency masker and the flanking bands, by introducing precursor flanking bands, and by introducing following flanking bands. Due to the large spacing (and the relatively low presentation levels), only across-channel processes contributed to CMR. In contrast, using narrowly spaced flanking bands with 1/6-octave spacing (similar to the conditions with close frequency spacings in Experiment 2), CMR was not affected by any of the stimulus manipulations. It was therefore suggested that (i) the within-channel mechanisms in CMR might be peripheral (brainstem level or below) in nature and therefore not susceptible to manipulation by auditory grouping constraints, and that (ii) the "slower" across-channel processing that is strongly dependent on auditory grouping constraints, might be of more central origin (see Chapter 4 and Dau et al., 2009).The model

investigated in the present study is not able to identify or extract auditory objects based on comodulation. A more advanced version of the model might apply basic concepts of computational auditory scene analysis (Bregman et al., 1990), where the EC-process would be switched on or off depending on the current spectro-temporal acoustical context.

2.7.2 Correlation with physiological CMR results

Even though a large number of studies have investigated CMR from a psychophysical perspective, little is known of its underlying physiological mechanisms (see, e.g., Verhey et al., 2003, for a review). A few studies have addressed physiological mechanisms of across-frequency processing by estimating signal-detection thresholds from the recordings of single- and multi-unit recordings in CMR-like paradigms. Several stages of processing along the auditory pathway were considered. Some studies intended to investigate across-channel processing but actually studied mostly within-channel cues due to the specific choice of the stimuli (e.g., Mott et al., 1990). Nelken et al. (1999) investigated the response of neurons in the primary auditory cortex to noise of varying bandwidth. They found a correlate for CMR in the band-widening paradigm in the disruption of the neurons' envelope following response. For most of the neurons in the population, the envelope locking was degraded by the addition of the pure tone signal. Using statistical criteria to estimate signal detection threshold, Nelken et al. (1999) demonstrated that the suppression of envelope locking lowers the detection thresholds for the single tones when comparing the responses of modulated versus unmodulated noise bands.

When considering true across-channel CMR, two possible correlates have been discussed recently. In the primary auditory cortex (of the cat), Rotman et al. (2001) in another study used a stimulus centered on the best frequency of the neuron and added two flanking bands equally spaced at either side of the best frequency. They showed that a single unit in the auditory cortex can demonstrate a response consistent with CMR in the flanking band paradigm. The correlate of CMR was again found as a disruption of the envelope following response. Thus, it appears that CMR is coded at a relatively late stage of auditory processing (in the primary auditory cortex) which appears conceptually compatible with the psychophysical findings on grouping

constraints on CMR. Their finding of very similar correlates for CMR in the two stimulus paradigms seems to differ from the modeling analysis discussed in the present study, which suggests very different mechanisms for the two processes.

A second physiological correlate of across-channel CMR has been suggested to be wide-band inhibition at brainstem level (e.g., Pressnitzer et al., 2001; Meddis et al., 2002). Here, it has been suggested, based on physiological experiments with the flanking-band paradigm with deterministic maskers, that cochlear nucleus onset units provide wide-band inhibition at the level of the brainstem onto narrow-band units in the ventral cochlear nucleus, and that this wide-band inhibition could provide a possible physiological basis for a potential EC model of CMR (for details about hypothetical neural circuits underlying CMR in the cochlear nucleus, see Pressnitzer et al., 2001; Verhey et al., 2003). A problem with such a neural correlate at the level of the brainstem might be the perceptual findings in the context of auditory grouping which make it unlikely that across-channel CMR can be accounted for by processing in the auditory brainstem and below.

A very promising way to fully understand the physiological mechanisms underlying CMR might be to study the correlation between neural responses and performance in the same species. Such an investigation was undertaken by Langemann and Klump (2001) and Nieder and Klump (2001) using the starling. Nieder and Klump (2001) investigated across-channel CMR with the flanking band paradigm, but used 100-Hz wide on-frequency and flanking bands amplitude modulated at 10 Hz. They showed that neural detection threshold was lowest when the probe tone was positioned in a dip of the masker envelope. They concluded that their multi-unit recordings in the auditory forebrain of the starling can be compared to the behavioral results in the same species. It would be interesting to specifically study the three basic paradigms of the present study in the same animal model both behaviorally and physiologically to learn more about the potential correlates of the different mechanisms underlying CMR.

2.7.3 Limitations of the current modeling approach

This study proposed an auditory signal processing model that accounts both for within-channel and across-channel processing in CMR. However, only three basic

experiments were considered in order to evaluate the model and to discuss the main principles of auditory processing underlying CMR – in the framework of the model. A number of experimental conditions have been investigated in previous CMR studies, which have not been considered directly in the present study. These studies investigated in much more detail effects of signal frequency (e.g., Schooneveldt and Moore, 1987), masker spectral width (Haggard et al., 1990; Hall et al., 1990) and masker spectral level (Moore and Shailer, 1991; Bacon et al., 1997; Cohen, 1991; Hall, 1986; McFadden, 1986), the influence of the envelope statistic of the masker modulator (e.g., Eddins and Wright, 1994; Grose and Hall, 1989; Moore et al., 1990; Hicks and Bacon, 1995), the effect of modulation frequency and modulation depth (Carlyon and Stubbs, 1989; Hall et al., 1996; Lee and Bacon, 1997; Bacon et al., 1997; Verhey et al., 1999; Eddins, 2001), effects of flanking band number and flanking band level (e.g., Hatch et al., 1995; Schooneveldt and Moore, 1987) and other effects. The current version of the model does not include a nonlinear peripheral filtering stage and therefore cannot account for level-dependent cochlear compression and effects associated with it such as level-dependent frequency tuning and suppression. While suppression does not seem to play a role in CMR with the level combinations in the present study (Haggard et al., 1985; Schooneveldt and Moore, 1987), effects of frequency selectivity certainly do, as was also shown in the present study. However, while corresponding modifications will change the details of the modeling outcomes, the main principles and implications discussed in the present study are expected to remain valid.

A further potential generalization of the model would be to include effects of dichotic presentation of flanker bands on CMR. The size of across-ear effects on CMR (2-3 dB) typically corresponds to that found in monaural across-channel CMR with one flanking band. The idea would be to apply the "central" EC mechanism to the stimuli after consideration of the inputs coming from the two ears. A binaural signal processing model based on the model by Breebaart et al. (2001a,b,c) but including a modulation filterbank stage is currently under development in order to process static as well as dynamic binaural stimuli (Thompson and Dau, 2008).

2.8 Summary and Conclusion

- A monaural auditory processing model was proposed that accounts for comodulation masking release (CMR) obtained in perceptual listening tests. The model distinguishes between contributions to CMR from within-channel processing and those resulting from explicit across-channel processing. For the across-channel process, an equalization-cancellation stage was assumed, conceptually motivated by models on binaural processing.

- The model accounts for the main findings in three critical experiments of CMR: (i) CMR with widely spaced flanking bands (where only across-channel processing contributes), (ii) CMR with one flanking band varying in frequency (where within-channel processing dominates at small separations while across-channel processing takes over at large separations), and (iii) CMR obtained in the classical band-widening experiment (where within-channel processing can never be eliminated).

- The simulation results support the earlier hypothesis that (at least) two different processes can contribute to CMR. The within-channel contributions can be as much as 15 dB and is caused by changes of the envelope statistics of the stimulus due to the addition of the signal to the (comodulated) masker – at the output of the auditory filter tuned to the signal frequency. The across-channel process is robust but small (about 2-4 dB) and only observable at small flanker bandwidths (below about 50 Hz).

- Specifically, in the classical band-widening experiment, which originally was used to define CMR as an across-channel process, the simulation results suggest that across-channel processing is not effective, not even at the largest noise bandwidth considered (2000 Hz) where several auditory filters are excited. CMR in this type of stimulus paradigm is dominated by within-channel processes.

- The current implementation of the model does not include a nonlinear, level-dependent cochlear filtering stage which limits its applicability in some of the

experimental conditions tested in previous CMR studies. The effect of a level-dependent frequency selectivity was investigated in one of the experiments of the present study using gammachirp instead of gammatone filters. A more complete implementation in the framework of the whole model is currently under investigation. Overall, the proposed model might provide an interesting framework for the analysis of fluctuating sounds in the auditory system.

3

Modeling comodulation masking release: Towards a unifying approach

3.1 Abstract

This chapter [1] presents an unified auditory processing model that accounts for the perceptual phenomenon of comodulation masking release (CMR) in different experimental conditions. It is based on the model of chapter 2 and additionally includes a compressive non-linear filtering stage acting instantaneously. The stage is supposed to mimic a more realistic filtering behavior of the basilar membrane. The model was evaluated in five experimental conditions: (i) CMR in a bandwidening type of paradigm as a function of masker spectrum level, (ii) CMR with four flanking bands varying in overall level in order to investigate level dependency of the across-channel mechanism, (iii) CMR with one flanking band varying in frequency and level relative to the on-frequency masker in order to investigate suppression based CMR, (iv) CMR with one flanking band varying in frequency in order to study the role of peripheral compression in conditions where CMR is dominated by within-channel processing and (v) CMR in a flanking band paradigm with a varying number of flanking bands. The simulations support the hypothesis that at least three different mechanism contribute to overall CMR. First, a within-channel process (as large as 15 dB) based on temporal beating cues between the signal and masking components. Second, a within-channel process that is based on suppression of the on-frequency masker's envelope caused by flanking bands. Dependent on the level of the masker the release can be as large as 9 dB. Finally,a "true" across-channel process that is

[1] *This chapter was submitted to JASA as Piechowiak et al. (2009)*

robust across levels but small (about 2-4 dB) and only observable at small masker bandwidths. Overall, the proposed model might provide an interesting framework for the analysis of fluctuating sounds in the auditory system.

3.2 Introduction

Many aspects of masking can be accounted for by the power spectrum model of masking (Fletcher, 1940) assuming that a listener's ability to detect a signal in a noise background is determined by the amount of the noise that passes through a single auditory filter with a center frequency close to that of the signal. That filter passes the signal but removes most of the noise and the threshold for the signal is assumed to correspond to a certain signal-to-noise ratio at the output of the filter. In the power spectrum model, the stimuli are represented by their long-term spectra, i.e., the relative phases of the components and the short-term fluctuations in the masker are ignored. Thus, the models fails, for example, in a condition when the masker has a coherent pattern of amplitude modulations across frequency. In such a condition, signal detection thresholds can be substantially lower than in the case of a noise of identical power that has uncorrelated amplitude modulations across frequency. This effect was named comodulation masking release (CMR) and was first demonstrated by Hall et al. (1984). Everyday sounds like speech exhibit comodulations and it is generally assumed that the effective processing of such envelope fluctuations across frequency represents a powerful survival strategy in the natural world (e.g., Klump and Langemann, 1995).

CMR has been studied using several experimental paradigms. In one class of CMR experiments, using the so-called bandwidening paradigm, the detection of a tone is measured as a function of the bandwidth of a noise masker (Hall et al., 1984). Two types of maskers have been used: One is a random noise with irregular fluctuations in amplitude that are independent in different frequency regions. The other is a random noise which was amplitude modulated using a low-pass filtered noise as a modulator. This modulation results in slow fluctuations in the amplitude of the noise that are the same in different frequency regions. The difference in detection threshold between them defines the amount of CMR in this paradigm. For the random noise, the detection threshold increases as the masker bandwidth increases up to about the critical bandwidth at that frequency and then remains constant, as expected from the power spectrum model (Fletcher, 1940; Patterson and Moore, 1986). For the modulated noise, the threshold pattern is quite different. Here, the threshold decreases

as the bandwidth decreases beyond about 2 kHz (for a signal frequency of 2 kHz); thus, adding more noise to the masker makes the signal easier to detect. The amount of CMR between these two conditions can be as large as 15 dB. The fact, that the decrease in threshold with increasing bandwidth only occurs with the modulated noise indicates that fluctuations in the masker are critical and that the fluctuations need to be correlated across frequency bands.

In another class of experiments, narrow-band noise maskers with inherently slow amplitude fluctuations have been used. One band, the on-frequency masker, is typically centered around the signal frequency, and one or several additional bands are centered remotely from the signal frequency. When the envelopes of the bands are correlated, the detection thresholds are usually lower than when they are random or when only the on-frequency masker is presented. The masking release observed using this paradigm ranges from 3 up to 15 dB (see Chapter 2 and Schooneveldt and Moore, 1987), depending on the spectral location of the signal and flanking bands and the number of the flanking bands.

It has also been shown that CMR is susceptible to level. Increasing the masker level typically increases the amount of CMR. Moore and Shailer (1991) investigated the level dependency of CMR using a single bandpass filtered noise as the masker. The signal frequency was 2000 Hz and the masker was centered around the signal frequency. CMR was measured as a function of the spectrum level of the noise masker and was found to increase with increasing spectrum level. For example, for a 3200-Hz wide noise masker, a spectrum level of 50 dB led to a 6 dB larger CMR than a noise spectrum level of 10 dB. When narrow-band noise maskers were used that were widely separated in frequency, the amount of CMR depended on the level difference between on-frequency and flanking band. For example, Ernst and Verhey (2006) showed that a level difference of 60 dB between the high-level flanking band and the lower-level on-frequency masker band can produce a CMR effect of 10 dB, considering the condition with only the on-frequency masker present as the reference. In their study, the flanking band was presented 3 octaves below the on-frequency band.

Even though CMR has been investigated in a number of studies, the underlying mechanisms are still not clear. Several studies have proposed that within-channel cues, i.e., information from only the one auditory channel tuned to the signal frequency,

can account for a considerable part of the effect in some conditions, which means that within-channel processing can lead to an overestimation of "true" across-channel CMR (e.g., Schooneveldt and Moore, 1987). This was supported by simulations of data from the bandwidening experiment, using a modulation filterbank analysis of the stimuli at the output of the auditory filter tuned to the signal frequency (Verhey et al., 1999). Additionally, for the CMR experiments using flanking bands, McFadden (1986) pointed out that it is imprecise to assume that one channel is receiving only the on-frequency band plus signal and another channel is receiving only the flanking band. Often, the two bands will be incompletely resolved. When this happens, the resulting waveform may contain envelope resulting from beats between the carrier frequencies of the on-frequency and the flanker bands. These beats can facilitate signal detection without across-channel comparisons being involved.

Peripheral suppression might also contribute to within-channel CMR. Suppression has been found at the level of the basilar membrane (BM), (e.g. Rhode and Robles, 1974; Ruggero et al., 1992) and in auditory-nerve firing patterns (e.g. Sachs and Kiang, 1968; Arthur et al., 1971). In terms of two-tone suppression, for example, the firing rate in response to a tone is reduced in the presence of a second tone with appropriate level, depending on the frequency separation between the two tones. The occurrence of suppression has been linked to the compressive nonlinearity in the processing on the BM. Hall et al. (1984) measured CMR with one flanking band for different spectral separations below or above the on-frequency band. They found the same amount of CMR regardless of whether the flanking band was presented below or above the signal frequency. However, the amount of suppression has been found to be larger on the high-frequency side than on the low-requency side at medium sound pressure levels (e.g. Sachs and Kiang, 1968; Houtgast, 1972). Thus, the CMR result in Hall et al. (1984) did not seem to be consistent with the "existence region" of suppression and the influence of suppression on CMR was discarded. Schooneveldt and Moore (1987) came to the same conclusion when investigating CMR in a similar paradigm.

Meddis et al. (2001) used a dual-resonance-nonlinear (DRNL) filter stage that contains a compressive nonlinearity to simulate two-tone suppression. In this framework, in the nonlinear path of the DRNL, the target tone and the suppressor tone

are filtered by a first bandpass (gammatone) filter, compressed by the compressive nonlinearity, and then filtered by a second bandpass (gammatone) filter. The second filter eliminates most of the suppressor's energy and provides a suppressed target tone at its output. If only the target tone alone is considered, its level is not large enough to be affected by the compressive nonlinearity. Ernst and Verhey (2006) used a single DRNL filter combined with a temporal-window detector ((Plack et al., 2002)) to simulate CMR in a flanking-band paradigm where the flanking band level was higher than or equal to that of the on-frequency masker band, and the on-frequency masker alone was taken as the reference. The simulated CMR was shown to increase with increasing level of the flanking band. They showed that the compressive nonlinearity, at least qualitatively, can account for CMR in this paradigm: When comodulated noise bands are used and their overall energy is large enough to fall in the compressive region of the BM input-output function, their modulations add constructively. The compressive nonlinearity reduces their modulation depths and thus also leads to a reduced modulation depth of the on-frequency masker band at the output of the nonlinear path. This results in a larger signal-to-noise ratio and thus a lower signal detection threshold in the framework of the model. This is not the case when only the on-frequency masker is considered since the masker alone does not have enough energy to reach the compressive region and no release of masking is simulated.

In addition to within-channel cues, several across-channel mechanisms have been proposed to contribute to CMR. One hypothesis is based on the assumption that the addition of the signal to the on-frequency masker band leads to a change in the modulation depth in the auditory filter centered at the signal frequency. By comparing this modulation depth to that of other auditory filters for which the modulation depth is unaltered, subjects would increase their sensitivity to the presence of the signal (Hall, 1986). A different explanation for CMR was proposed by Buus (1985), who suggested that the comodulated flanker band(s) provide valuable information about the moments at which the masker level has a relatively low energy. By attributing more weights to these valleys in the masker, the effective signal-to-noise ratio increases and detection improves. This mechanism was called "listening in the valleys". Also proposed by Buus (1985) was an equalization-cancellation (EC) mechanism, originally introduced by Durlach (1963), to account for various binaural masking release data. According to

this mechanism, the envelope of the masker and flanking band are first equalized and then subtracted. In Chapter 2 an EC circuit was implemented into an extended version of the modulation filterbank model by Dau et al. (1997a), whereby the EC processing was assumed to take place after modulation filtering at a central stage of processing. This model was able to account for various aspects in the data associated with across-channel CMR. However, the model does not include a nonlinear, level-dependent cochlear filtering stage which clearly limits its applicability in some experimental conditions.

The goal of the present study was to present and evaluate a unified auditory processing model that allows to quantify the (relative) contributions of across- versus within-channel processing in the various types of CMR conditions, and to analyze the effects of level-dependent nonlinear cochlear processing on CMR. The computational auditory signal processing and perception (CASP) model developed by Jepsen et al. (2008) was used as the framework. This model is based on the modulation filterbank model of Dau et al. (1997a) but contains a nonlinear cochlear stage, the DRNL filterbank Meddis et al. (2001) instead of the original gammatone filterbank (Patterson and Moore, 1986). Here, the CASP model was extended by an EC processing stage based on the model in Chapter 2 in order to integrate across-channel modulation processing into the model.

First, the structure of the generalized processing model will be described. This will then be tested in several CMR conditions considering (i) effects of masker level in the bandwidening paradigm, (ii) effects of overall and relative masker levels in the flanking-band paradigm, (iii) effects of spectral separation between masker and flanking bands and (iv) effects of the number of flanking bands. In a separate model analysis, the role of compression on CMR, the concept of a single- versus a multi-channel model analysis and, finally, the interaction between nonlinear cochlear processing and modulation filtering as well as the EC mechanism will be investigated. Finally, the implications of this work will be discussed.

3.3 Model

Figure 3.1 shows the structure of the proposed auditory processing model. It is based on the computational auditory signal processing (CASP) model developed by Jepsen et al. (2008), which itself originates from the modulation filterbank model of Dau et al. (1997a).

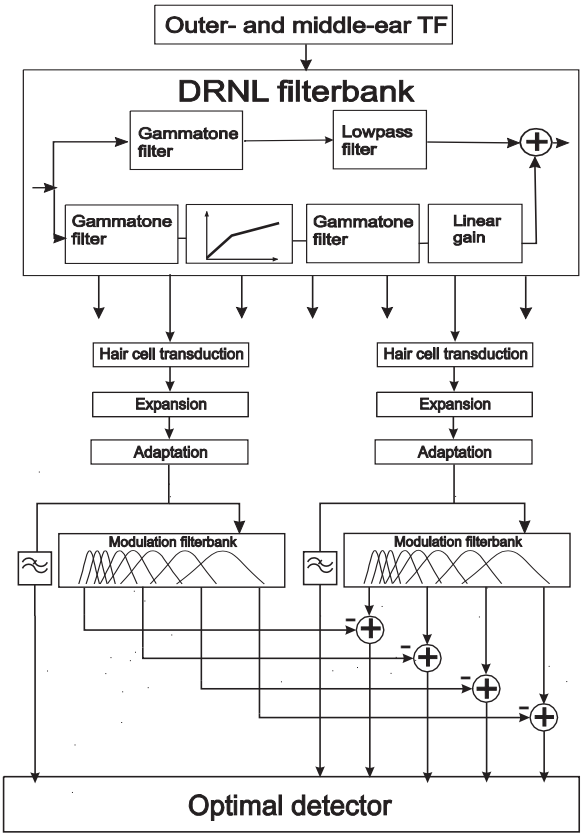

Figure 3.1: CASP model with EC mechanism. The EC mechanism is illustrated here exemplarily for two auditory filters

The first stage is a bandpass filter that represents the transformation of the input

stimuli through the outer and the middle ear. Its transfer function is given by a symmetric bandpass filter with a CF of 800 Hz and and a rolloff of 20 dB/oct and -20 dB/oct respectively. The output of this filter is the frequency dependent stapes velocity. In contrast to the original across-channel model configuration (see Chapter 2 and Dau et al., 1997a) the bandpass characteristic of the basilar membrane is simulated by a DRNL filterbank. This filter type was developed by Meddis et al. (2001) to introduce a level- and frequency-dependent compression and to account for peripheral filter shape in animals (e.g., Ruggero et al., 1997). The DRNL consists of two independent parallel processing pathways, one linear and the other one compressive non-linear. The linear path consists of a linear gain followed by a gammatone and a lowpass filter whereas in the non-linear pathway the compressive non-linearity is preceded by a gammatone filter and followed by a gammatone and a lowpass filter. The output of the overall filter represents the sum of the outputs of the nonlinear and linear part. The compressive non-linearity can be best characterized by its input/output (I/O) functions. The (I/O) function is nearly linear at or close to center frequency at low input levels and compressive (0.2 - 0.5 dB/dB) at medium-to-high input levels whereas it is linear (0.8 - 1 dB/dB) at frequencies below the center frequency.

The output of each peripheral filter is followed by a half-wave rectification and low-pass filtering at 1 kHz in order to roughly simulate the transformation of the mechanical BM oscillations into receptor potentials. The lowpass filtering preserves the temporal fine structure of the signal at low frequencies and extracts the envelope of the signal at high frequencies (e.g., Palmer and Russell, 1986). The hair-cell-transformed signal is squared in an expansion stage into an intensity-like quantity since experimental evidence show that the dependency between stimulus level and AN nerve fiber firing rates follows a square law (e.g. Yates et al., 1990). The simulation of AN nerve adaptation is performed by a series of non-linear adaptation loops (Dau et al., 1996) which introduces a logarithmic compression for stationary signals and an almost linear transformation for fast fluctuating stimuli. Regarding the transformation of envelope fluctuations, the adaptation loops transforms the amplitude modulation depth of input fluctuations with rates higher than about 2 Hz almost linearly. The output of the adaptation loops is further processed by a 150 Hz lowpass filter to

simulate a decrease in sensitivity for modulation processing (Ewert and Dau, 2000). This filter is followed by a modulation filterbank with a second order butterworth lowpass filter with the center frequency of 2.5 Hz as the lowest modulation filter. Modulation filters at and above 5 Hz are bandpass filters. The filters with 5 and 10 Hz center frequency have a constant bandwidth of 5 Hz. Modulation filters with higher center frequency have a quality factor of $Q = 2$. The highest modulation center frequency is taken as one-quarter of the maximal peripheral center frequency and maximally 1000 Hz (e.g., Langner, 1992). For modulation frequencies below 10 Hz, only the real part of the filter is processed and for center frequencies above 10 Hz, the Hilbert envelope of the filter is considered.

The across-channel process is the same as proposed in Chapter 2. This stage is conceptually close to the EC mechanism of Durlach's model (Durlach, 1960, 1963) for describing binaural masking level differences (BMLDs). However, while the EC mechanism in the original (binaural) models is essentially applied to the stimulus *waveforms*, and jitter is provided in the level and time domains in order to limit resolution in the model, the (monaural) EC process in the current model is applied at a later stage of auditory processing, and no additional limitations are introduced. In contrast to the original binaural EC model, it is assumed here that the limitations of performance are already included in the processing steps prior to the EC process. The across-channel processing within the model is assumed to occur at the output of all (bandpass) modulation channels tuned to frequencies at and above 5 Hz, which is the center frequency of the lowest modulation filter. The individual modulation filter outputs at the flanking bands are subtracted from the corresponding outputs at the on-frequency channel (the cancellation process). The outputs of the lowpass filters in the different peripheral channels remain unaffected. Usually more filters than one remote peripheral filter are considered. The activity of all remote peripheral filters is weighted according to their energy and then averaged. This averaged activity is subtracted from the on-frequency channel. Calculating the weighted sum can be considered as equalization process, since it equalizes the summed activity in the different flanking bands with regard to the on-frequency band. Weights are set to zero when the correlation of the output of peripheral channels neighboring the signal channel were less than 5% using a broadband noise as input and a gammatone filterbank as the

peripheral filtering stage. This is done in order to ensure statistical independence from the signal channel to the channels involved in the EC process.

In the present paper, the EC type process was generalized in the way that each peripheral channel was considered as a potential signal channel. This was achieved by applying the EC process on each single peripheral channel. Formally, the EC mechanism presented here is equivalent to a correlation analysis (Green, 1992; Richards, 1987) whereby the amount of CMR is determined by the change of correlation when the signal is added. In this sense, the EC mechanism leads to a de-correlation of the peripheral channels and thus to a higher sensitivity in detecting a target signal (see 2). The detection of a target signal takes place at the output of the EC mechanism assuming an optimal detector which is comparable to a "matched filtering" process (Green and Swets, 1966). A stored, normalized temporal representation of the signal, that is supposed to be detected, the template, is compared with the internal representation of the actual signal, the reference, by calculating the cross correlation between these two temporal patterns (Dau et al., 1996, 1997a).

The simulations in the following will be based on a multi-channel model in the sense that the integrated information across peripheral channels will be considered in the detection process. This is in contrast to a single-channel model where only the output of one peripheral filter channel is used for detection. The "complete" model contains the multi-channel (MC) nonlinear (NL) DRNL filterbank, the modulation-filterbank (MFB), and finally the EC-type (EC) stage. This model configuration is thus in the following also referred to as "MC_NL_MFB_EC". All simulations presented in section 3.4-3.8 will be based on this model version.

For analysis purposes, the results obtained with the complete model were compared with results obtained with modified versions (section 3.9) in order to address the (relative) contributions of different model stages to CMR. The following components were considered as key stages: the type of peripheral processing (linear (L) versus nonlinear (NL)), the number of peripheral channels considered in the detection process (single-channel (SC) versus multi-channel (MC)), the processing of modulations (modulation filterbank (MFB) versus lowpass filtering (LP)) as well as the influence of the EC process (on (EC) versus off).

The DRNL filterbank was used with parameters according to Jepsen et al. (2008).

The frequency range covered by the filterbank depended on the stimuli in the different experimental paradigms. This frequency range was chosen such that at least one peripheral filter was centered above and one below the frequency range covered by the stimuli. The center frequencies of the DRNL filters were chosen with a spacing according to the equivalent rectangular bandwidth spacing in the case of gammatone filters.

3.4 Experiment 1: Bandwidening paradigm

3.4.1 Rationale

In the bandwidening paradigm, CMR is measured as a function of the bandwidth of a masker centered on the target tone. In earlier studies (e.g., Schooneveldt and Moore, 1987; Verhey et al., 1999), the results from the bandwidening experiment have been explained mainly in terms of within-channel processing. Model simulations obtained with the across-channel model of Chapter 2 confirmed the dominance of within-channel cues in this paradigm. Here, the effect of non-linear multi-channel peripheral processing on CMR was investigated, whereby the masker spectrum level was varied and the masker bandwidth was kept fixed. Simulations were compared with own measured data.

3.4.2 Method

Apparatus and procedure

A three-interval, three-alternative forced choice procedure was used to measure detection thresholds. A one-up two-down procedure was used to estimate the 70.7% correct point on the psychometric function (Levitt (1971)). Subjects had to identify the one randomly chosen interval containing the signal. Subjects received visual feedback if the response was correct. The initial step size for the signal level was 8 dB and was halved after every reversal until the final step size of 1 dB was reached. The mean of the signal level at the last six reversals was calculated and regarded as the masked threshold value. For each stimulus configuration and subject, four masked

thresholds were measured and averaged to obtain final thresholds. All stimuli used in the following were generated and presented digitally by the AFC software package for MATLAB 7.2® and then transformed to the analogue domain by a 32-bit soundcard (RME) at a sampling rate of either 32 or 44.1 kHz. Measurements were performed in a sound attenuated double walled booth.

Subjects and stimuli

Four normal hearing subjects (three male, one female) between 24 and 32 years participated in this experiment. All subjects had prior experience in psychoacoustic experiments. The stimuli were similar to those used in Verhey et al. (1999). The signal was a 2000-Hz tone with a duration of 300 ms including 50 ms cosine ramps. The masker was a single noise band with a duration of 600 ms including 10 ms cosine ramps. The signal was temporally centered within the masker. The masker band was a 2000-Hz wide noise. It was either comodulated or unmodulated. In the comodulated case, a white noise was multiplied with a lowpass noise with a cutoff frequency of 50 Hz. The spectrum level of the noise band was -10, 10, 30, or 40 dB.

Simulation parameters

Peripheral channels in the range from 1 to 3 kHz were considered in the simulations. Simulated thresholds were determined using the same stimuli and procedure (3 AFC, 1-up 2-down) as in the measurements. The average of 10 repetitions for each condition was taken as the final simulated threshold.

3.4.3 Results

The left panel of Fig. 3.2 shows the measured data as a function of the masker spectrum level. In the random condition (circles), the relative signal threshold does not depend strongly on the masker level and lies between 20 and 25 dB relative to the masker spectrum level. In the comodulated condition (squares) signal threshold drops strongly between -10 and 10 dB masker level and then stays roughly constant for higher masker levels. CMR is determined as the signal threshold difference between the uncorrelated and comodulated condition. For the lowest spectrum level,

no CMR is observed. The reason for this is that the measured threshold approaches the absolute signal threshold. For a masker spectrum level of 10 dB, the threshold in the comodulated condition lies about 15 dB below the threshold in the uncorrelated condition. The amount of CMR is 7 dB for a masker spectrum level of 10 dB and 10 dB for a masker spectrum level of 40 dB which is mainly due to the increase of the signal threshold with level for the uncorrelated masker. The results here are similar to the results from Moore and Shailer (1991) although a longer signal duration was applied in the present Chapter. The results for the 30 dB masker spectrum level are consistent with Verhey et al. (1999) where a CMR of about 11 dB was found for a 2000-Hz masker bandwidth and a spectrum level of 30 dB.

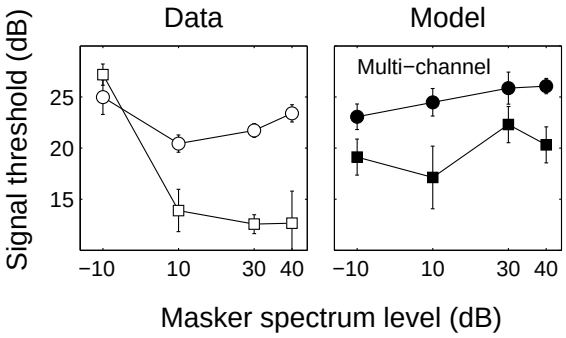

Figure 3.2: Signal thresholds as a function of the masker spectrum level when the masker band was uncorrelated (circles) and when it was correlated (squares). Left panel: Averaged measured data from four subjects. The right panel shows simulated data with the "MC_NL_MFB_EC" configuration.

The simulated results are shown in the right panel of Fig. 3.2. In the uncorrelated condition, the signal threshold increases slightly with increasing masker level as a result of the increasing auditory filter bandwidth with level. The thresholds are close to the measured data. However, for the comodulated condition, the simulated thresholds are above the measured ones (except for the lowest masker level condition); thus, the model clearly underestimates the amount of CMR. This observation is inconsistent with the simulations provided by Verhey et al. (1999) which showed large agreement with the experimental data for this masker bandwidth and a masker spectrum level of

30 dB. Verhey et al. (1999) presented single-channel simulations considering only one gammatone filter tuned to the signal frequency while the results from a multi-channel simulation were considered here. A detailed analysis of the effects of single- versus multi-channel and gammatone versus DRNL processing will be provided further below (section 3.9.2).

3.5 Experiment 2: Effects of masker level in the flanking-band paradigm

3.5.1 Rationale

The level dependency of CMR was considered using the flanking-band paradigm with octave spacing between the flankers. In this experiment, it can be expected that CMR is mainly determined by across-channel processing. At medium levels, a CMR effect of about 3-4 dB was found in this condition (see Chapter 2), and the data were successfully accounted for by the across-channel modulation filterbank model assuming a *level independent* EC process. Here, the question was whether the across-channel process needs to be level dependent or whether it is sufficient to assume a level independent stage as assumed in Chapter 2.

3.5.2 Method

Apparatus and procedure

A three-interval, three-alternative forced choice procedure was used to measure detection thresholds. A one-up two-down procedure was used to estimate the 70.7% correct point on the psychometric function (Levitt (1971)). Subjects had to identify the one randomly chosen interval containing the signal. Subjects received visual feedback if the response was correct. The initial step size for the signal level was 8 dB and was halved after every reversal until the final step size of 1 dB was reached. The mean of the signal level at the last six reversals was calculated and regarded as the masked threshold value. For each stimulus configuration and subject, four masked thresholds were measured and averaged to obtain final thresholds.

Subjects and stimuli

Four normal hearing subjects (four male) between 24 and 29 years participated in the experiment. All had prior experience in psychoacoustic experiments. The target signal was a 1000-Hz pure tone. The masker consisted of five bands of noise which were centered at 250, 500, 1000, 2000 and 4000 Hz covering a four octave range. Signal and masker had a duration of 187.5 ms including 20-ms raised-cosine ramps. Signal thresholds were measured as a function of the level of the masker band which was 35, 55 or 75 dB sound pressure level (SPL). The width of the masker bands was 25 Hz. The masker was generated in the time domain, transformed to the frequency domain by Fourier transformation and then restricted to the 25-Hz bandwidth. In the reference condition, the envelopes of the five bands were uncorrelated with each other. In the comodulated condition, the on-frequency masker was shifted to the center frequency of the flanking bands. In this way, the envelopes of the masker bands were fully correlated.

Simulation parameters

Peripheral channels in the range from 0.2 to 4.5 kHz were considered in the simulations. Simulated thresholds were determined using the same stimuli and procedure (3 AFC, 1-up 2-down) as in the measurements. The average of 10 repetitions for each condition was taken as the final simulated threshold.

3.5.3 Results

The results of the simulations are shown in Fig. 3.3. Signal thresholds are plotted relative to the masker level of the individual noise bands for the random (circles) and the comodulated condition (squares). The left panel shows the average measured thresholds for four subjects. There is a significant CMR effect of around 4 dB for all masker band levels [one-way analysis of variance (ANOVA): $F(1,26) = 50.43, p < 0.001$ for 35 dB, $F(1,26) = 48.31, p < 0.001$ for 55 dB and $F(1,26) = 36.82, p < 0.001$ for 75 dB]. The right panel of Fig. 3.3 shows the corresponding simulations. The model produces significant CMR for all noise masker levels [one-way analysis of variance (ANOVA): $F(1,26) = 35.58, p < 0.001$ for 35 dB, $F(1,26) = 15.55, p <$

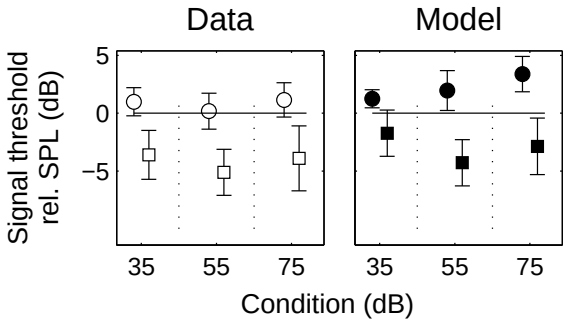

Figure 3.3: Left panel: Measured detection thresholds for a 1-kHz tone in the presence of five noise bands as a function of the SPL of each noise band. Circles and squares represent results for the uncorrelated and comodulated conditions, respectively. Right panel: Simulated signal thresholds

0.001 for 55 dB and $F(1, 26) = 16.92, p < 0.001$ for 75 dB]. The amount of CMR is in good agreement with that observed in the measurements. However, the simulated thresholds tend to increase with level in the reference condition, which is not observed in the data. Since the non-linear DRNL filter increases its filter bandwidth with increasing input level, the increase is due to more masker energy that is mapped into the signal channel (Jepsen et al., 2008). This causes the CMR at 35 dB masker level to be slightly smaller than for the other masker level. A one-way ANOVA test revealed that thresholds for the comodulated condition across masker levels are not significantly different from each other, both in the data ($F(2, 41) = 1.99, p = 0.16$) and in the simulations ($F(2, 29) = 3.5, p = 0.06$).

3.6 Experiment 3: Effect of relative masker levels of flanking and masker bands

3.6.1 Rationale

For large spectral separations between flanker and on-frequency masker band, CMR has been found in conditions where the flanking band had a much higher level than the on-frequency masker (e.g., Ernst and Verhey, 2006). In such conditions, CMR has

been associated with effects of suppression where the fluctuations in the on-frequency band might be effectively reduced due to the coherent fluctuations in the flanker band. It was shown earlier by Ernst and Verhey (2006) that the DRNL model can, in principle, account for CMR using the temporal-window modeling framework. Here, it was tested to what extent the model framework presented here can successfully account for these CMR data. Simulations were compared with the experimental data of Ernst and Verhey (2006).

3.6.2 Method

Stimuli

The target signal was a 2000-Hz pure tone. The on-frequency masker was a narrow band of noise centered at the signal frequency. The flanking band was created was created by multiplying a sinusoidal carrier with a 10-Hz wide lowpass noise resulting in 20-Hz wide noise bands. When the noise bands were comodulated the same lowpass noise was used for multiplication. 50-ms long raised-cosine ramps were applied to the bands. In the reference condition, signal threshold was determined when only the on-frequency masker with a level of 20 dB SPL was presented. In the comodulated condition, the flanking band was added with center frequencies of 250, 500, 1000 and 3031 Hz.

3.6.3 Simulation parameters

Peripheral channels in the range from 0.2 to 4 kHz were considered in the simulations. Simulated thresholds were determined using the same stimuli and procedure as in the measurements. The average of 10 repetitions for each condition was taken as the final simulated threshold.

3.6.4 Results

Figure 3.4 shows the measured amount of measured CMR (left panel), replotted from Ernst and Verhey (2006), whereby CMR reflects the threshold difference between the reference condition and the comodulated condition. CMR is plotted as a function of

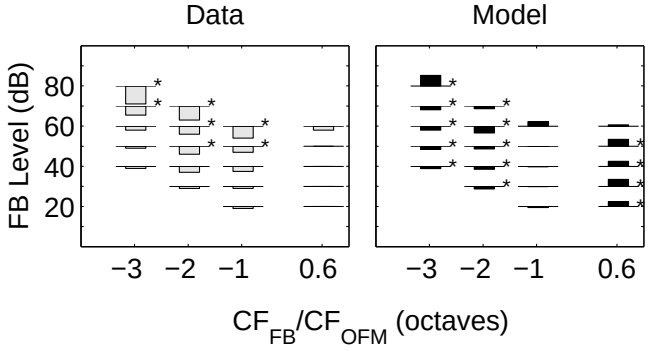

Figure 3.4: Barplot denoting the amount of CMR in a flanking band paradigm. The CMR is plotted as a function of spectral position of the flanking band relative to the signal frequency, shown on the abscissa and the level of the flanking band, shown on the ordinate. A release from masking, a positive CMR, is indicated by a bar pointing downwards. Negative CMR by a bar pointing upwards. Left panel: Measured data taken from Ernst and Verhey (2006). Right panel: simulated CMR with the auditory perception model.

the spectral position of the flanking band relative to the on-signal frequency band, indicated on the abscissa, and the level of the flanking band, indicated on the ordinate. Barplots are used to represent the amount of CMR. A release from masking, i.e. a positive CMR, is indicated by a bar pointing downwards. Negative CMR is indicated by a bar pointing upwards. Significant releases from masking are indicated by a star at the top right position of the barplots. The largest masking release of around 9 dB was found at a flanking band frequency of 250 Hz (-3 octaves) for a masker level of 80 dB. Ernst and Verhey (2006) showed that no CMR was observed when the flanking band frequency was placed above the signal frequency. For relative flanking band frequencies of -3 respectively -1 octaves, CMR was observed for the two highest levels. At 500 Hz, the three highest levels of the flanking band produced significant amount of CMR. Generally, CMR was found to increase towards lower flanking band frequencies Ernst and Verhey (2006).

The right panel of Fig. 3.4 shows corresponding simulated thresholds. There is no simulated CMR for all levels at relative flanking band frequencies of -3 and -2 octaves and for the four lowest levels of 0.6 octaves relative flanking band frequency. For the flanking band presented at -1 octave, no CMR is simulated at any level. Some

major deviations between the simulations and the data can be observed. Particularly in the conditions where the largest CMR was measured, i.e., at the highest levels for the flanking band frequency of -3, -2 and -1 octaves, the model clearly underestimates the amount of CMR. For example, for a flanking band frequency of -3 octaves and a level of 80 dB SPL, even a negative CMR effect is simulated in contrast to the measurements which showed a CMR effect of approx. 9 dB. The simulations here are in contrast to the simulations by Ernst and Verhey (2006) who found a better agreement with the measurements. Their model differed in three main aspects from the model used here. First, for the compressive nonlinearity, a different slope of compression was used. Second, a sliding temporal window (Oxenham and Moore, 1994) was used for detection instead of an "optimal detector" device. Third, the temporal window corresponds to a modulation lowpass filtering while the optimal detector is applied to the output of the modulation filterbank channels in the model considered here. In the analysis section further below, it will be demonstrated that mainly the slope of the assumed compression in the peripheral filtering stage affects the outcome of the simulations.

3.7 Experiment 4: Effects of spectral separation between masker and flanking band

3.7.1 Rationale

In this experiment, the transition between within-channel and across-channel CMR was investigated for a given masker level. A flanking-band paradigm was used and the spectral separation of the on-frequency and flanking band was varied. For small spectral separations, within-channel cues in the form of beating exists which become less prominent with increasing spectral distances where across-channel processing comes into play. The model introduced in Chapter 2 that assumes a gammatone filterbank could not account for the exact shape of the masking curves. Here, the effect of peripheral compression and level-dependent filter bandwidth on CMR in this experimental condition were investigated. The simulations were compared to the data of Chapter 2.

3.7.2 Method

Stimuli

The signal was a 2-kHz tone that was masked by two 25-Hz wide narrow bands of noise. The on-frequency masker was centered at the signal frequency and the flanking band was centered at 1000, 1400, 1800, 1900, 2100, 2200, 2600 and 3000 Hz. The masker was either random or comodulated. Comodulation was achieved by multiplying two sinusoidal carriers with the same low-pass noise with a cutoff frequency of 12.5 Hz. In the random situation each sinusoidal carrier was multiplied with a different low-pass noise.

Simulation parameters

The peripheral channels in the range from 0.9 to 3.5 kHz were considered in the simulations. The simulated thresholds were determined using the same stimuli and procedure as in the measurements. The average of 10 repetitions for each condition was taken as the final simulated threshold.

3.7.3 Results

The left panel of Fig. 3.5 shows the measured data from Fig. 2.6 in Chapter 2. Signal thresholds are plotted as a function of the ratio between flanking-band and signal frequency. Squares denote thresholds when the masker bands are comodulated and circles when they are uncorrelated. For small spectral separations between on-frequency and flanking band (with ratios between 0.9 and 1.1), CMR amounts to 12-14 dB. For larger spectral separations, the data show an asymmetry. CMR amounts to 4-5 dB for small flanking-band frequencies and 2-3 dB for high flanking-band frequencies.

The right panel shows the corresponding simulations. The model predicts slightly elevated overall thresholds as seen in previous sections. For small spectral separations between on-frequency and flanking band, the model predicts a large amount of CMR that corresponds to that found in the data. CMR is slightly underestimated at the (relative) spectral separations of 1.05 and 1.1 and slightly overestimated (by around

2 dB) for the separation ratios 0.9 and 0.95, due to elevated signal thresholds in the uncorrelated masker band condition. However, the models accounts nicely for the transition between the conditions that can be associated with within-channel processing and those associated with across-channel processing. In particular, the model accounts for the asymmetry in the data. This was *not* found in Chapter 2.Thus, the assumption of a more realistic peripheral processing stage has led to clearly better simulations. A detailed analysis of of the effects of fast-acting compression and level-dependent frequency tuning will be provided in section 3.9.1.

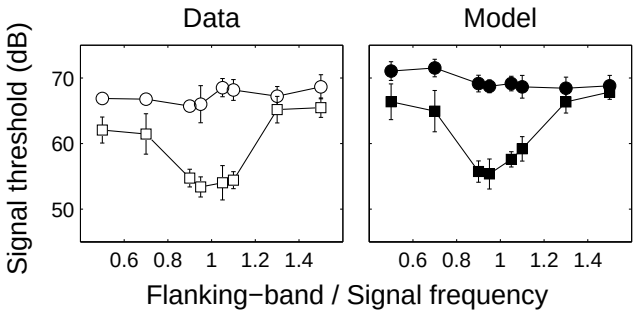

Figure 3.5: Left panel: Detection thresholds for the 2-kHz tone in the presence of two noise bands as the function of the spectral distance to the tone. Circles and squares represent results for the random (denoted by circles) and comodulated conditions (denoted by squares), respectively. Right panel: Simulated thresholds for a multichannel model configuration when a range of filter channels with the EC process turned on was considered.

3.8 Experiment 5: Effect of number of masker bands in different spectral configurations

3.8.1 Rationale

In this experiment, the influence of the number of masking bands on the amount of CMR was investigated. It remained unclear from previous studies investigating CMR as a function of the number of flanking bands (Hatch et al., 1995; McFadden, 1987) to what extent CMR resulted from within- versus across-channel processing. In order

to separate between these processes in this type of experiment, two conditions were considered here: In the first condition, the masking bands were placed close to each other to emphasize within-channel cues, according to the study of Hall et al. (1990). In the second condition, the flanking bands were broadly separated from each other to examine the influence on the number of masking components on across-channel cues. Own data were collected in this second condition.

3.8.2 Method

Apparatus and procedure

A three-interval, three-alternative forced choice procedure was used to measure detection thresholds. A one-up two-down procedure was used to estimate the 70.7% correct point on the psychometric function (Levitt (1971)). Subjects had to identify the one randomly chosen interval containing the signal. Subjects received visual feedback if the response was correct. The initial step size for the signal level was 4 dB and was halved after every reversal until the final step size of 1 dB was reached. The mean of the signal level at the last six reversals was calculated and regarded as the masked threshold value. For each stimulus configuration and subject, four masked thresholds were measured and averaged to obtain final thresholds.

Stimuli

The stimuli for the narrowband situation were the same as in Hall et al. (1990). The signal was a 700-Hz tone with a duration of 400 ms windowed with 50 ms long ramps. In the simplest case, the masker consisted only of a single 20-Hz wide noise band centered at the signal frequency (on-frequency band). The threshold from this condition served as the reference. Six conditions were further considered in which different configurations of flanking bands were presented in addition to the on-frequency band. All flanking bands were 20 Hz wide and were separated by 100 Hz. Two configurations with flanking bands only at the low-frequency side were used with: (i) one flanking band centered at 600, and (ii) three flanking bands centered at 400, 500 and 600 Hz. Two configurations with flanking bands only at the high-frequency side were used: (iii) one flanking band centered at 800 Hz, and (iv) three flanking bands at

800, 900 and 1000 Hz. Furthermore, two spectrally symmetric configurations with (v) two flanking bands centered at 600 and 800, and (vi) six flanking bands at 400, 500, 600, 800, 900 and 1000 Hz. The comodulated noises were generated by multiplication of a lowpass noise with a cutoff-frequency of 10 Hz with tones at 400, 500, 600, 700, 800, 900 and 1000 Hz, respectively. The masker bands had the same duration and ramps as the signal. Each masker band had a spectrum level of 50 dB resulting in 63 dB SPL for each band.

In the broadband condition, the signal was a 1000-Hz tone. Again, in the reference condition, only the signal and the on-frequency masker were presented. CMR was obtained in six configurations as in the narrowband condition. Here, a one-octave spacing between the flankers was used. The distribution of the flanking bands was as follows: (i) one flanking band centered at 500 Hz; (ii) three flanking bands centered at 125, 250, 500 Hz; (iii) one flanking band centered at 2000 Hz; (iv) three flanking bands centered at 2000, 4000 and 8000 Hz; (v) two flanking bands centered at 500 and 2000 Hz; and (vi) six flanking bands centered at 125, 250, 500, 2000, 4000 and 8000 Hz. The noise bands were 20-Hz-wide Gaussian noises. A broadband Gaussian noise was fourier-transformed and the coefficients outside the desired cutoff frequencies were set to zero. Finally, the band was transformed back to the time domain. Each masker band was presented at a sensation level of 60 dB. Sensation levels were derived by measuring the absolute threshold at the center frequencies of the flanking bands for all subjects individually. The resulting average sound pressure level of the bands were 64.6 dB SPL for the 125 Hz flanker, 69.7 dB SPL for the 8000 Hz flanker and 60 dB SPL for the remaining bands, respectively.

Simulation parameters

The peripheral channels in the range from 0.1 to 1.3 kHz were considered in the simulations. The simulated thresholds were determined using the same stimuli and procedure as in the measurements. The average of 10 repetitions for each condition was taken as the final simulated threshold.

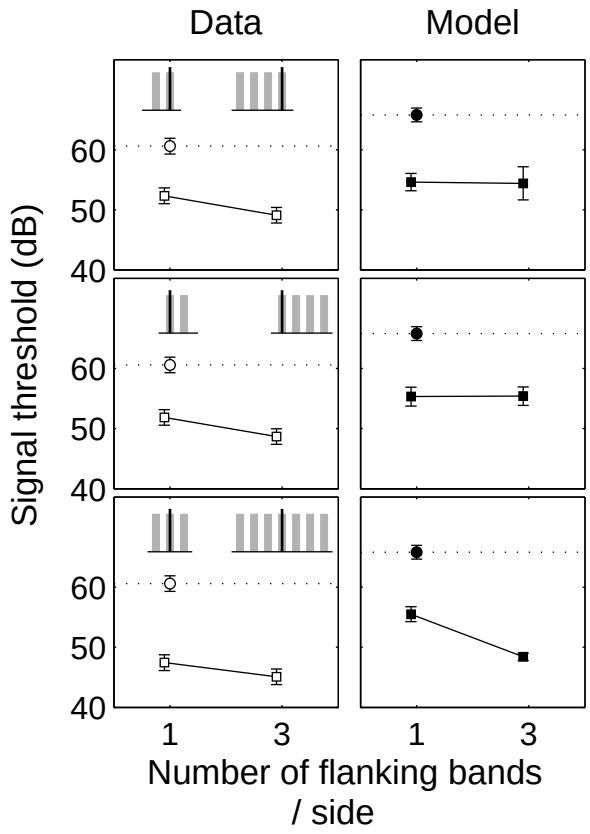

Figure 3.6: Measured and simulated signal thresholds as a function of the number and spectral location of comodulated flanking bands. The signal had a frequency of 700 Hz. Masking bands had a mutual separation of 100 Hz. Open squares denote measured, solid squares simulated thresholds. The dotted line indicates the signal threshold in the reference when only the on-frequency masker was presented. Left panels: Measured data and stimuli sketches. The gray blocks denote masking bands and the solid black line the signal. Right panels: Simulated data with the overall model.

3.8.3 Results

Figure 3.6 shows results for the narrowband condition. The left panels show measured signal thresholds. Open squares denote measured signal thresholds in the comodulated conditions. The circle shows the reference threshold. The spectral configurations of

the flanking bands are indicated by the sketches in the upper part of each panel. The vertical line indicates the signal frequency centered in the on-frequency masker. The upper panel shows the results for the flanking bands placed below the on-frequency band. CMR amounts to 8 dB when one flanking band is presented below the on-frequency band and increases to 10 dB when two bands were added. The middle panel shows the results with the flanking bands placed above the on-frequency masker. Also here, CMR increases from 8 dB for the condition with one flanking band to about 10 dB for the condition with three flanking bands. The bottom panel shows the results for the flanking bands placed symmetrically around the on-frequency band. The CMR amounts to 12 dB for two symmetrical flanking bands and 15 dB for six bands. This large amount of CMR is in line with earlier studies investigating CMR in similar conditions (e.g., Schooneveldt and Moore, 1987; Hall et al., 1984). The data suggest that the flanking bands located closest to the signal frequency contribute more to the observed overall CMR than additionally added flanking bands.

The right panel of Fig. 3.6 shows the corresponding simulations. The model generally slightly overestimates thresholds but the amount of CMR (7-10 dB) is similar to that found in the data. In the asymmetric conditions (upper two panels) no decrease in signal threshold was observed in the model when the number of flanking bands was increased from one to three. Only for the symmetrical condition, an increase of simulated CMR with increasing number of flankers was found, whereby the effect is in this case larger than in the data.

Figure 3.7 shows the results for the broadband condition where the masking bands had a spectral separation of an octave. Again, the left-hand column shows measured data and the right-hand column simulated data and the stimuli configurations as sketches. Circles denote signal thresholds when random masking bands were used. When the flanking bands were placed below the on-frequency masker (upper left-hand panel), the CMR amounted to 3 dB for when only one flanking band was used and when three bands were used. A two-way ANOVA (Random bands/comodulated bands: $F(1,12) = 69.18, p < 0.01$, number of flanking bands: $F(1,12) = 0.96, p = 0.35$, interaction of the two variables: $F(1,12) = 0.4, p = 0.54$) for this condition reveals the relation of envelopes (random or comodulated) as the only source of masking release.

In the case when only one flanking band was placed above the signal frequency no CMR was measured. This is in contradiction to findings from Schooneveldt and Moore (e.g. 1987); Hall et al. (e.g. 1984); McFadden (e.g. 1987) where in a comparable condition a CMR of around 3 dB was obtained. When three bands were presented a CMR of 3 dB seems to be measured. A two-way ANOVA (random bands/comodulated bands: $F(1,12) = 15.53, p < 0.01$, number of flanking bands: $F(1,12) = 8.41, p = 0.013$, interaction of the two variables $F(1,12) = 17.71, p < 0.01$) suggests a significant release of masking and also a dependency between the bands envelope relation and the number of band, i.e. more CMR with more bands. However, since the outcome for a single flanking band contradicts previous studies the dependency here has to be seen critical.

When symmetrically placed flanking bands (on an octave scale) were used (right-hand bottom panel) CMR seems to increase from 3 dB for two bands to 5 dB for six flanking bands (two-way ANOVA, random/comodulated masking bands $F(1,12) = 86.75, p < 0.01$, number of flanking bands $F(1,12) = 7.81, p = 0.016$, interaction $F(1,12) = 4.18, p = 0.06$). These values indicate that the only source of CMR here is the relation of the envelopes (random or comodulated) but not the number of flanker bands and there is no evidence of a synergistic (interaction) effect of the two. The statistical analysis leads to the conclusion that CMR remains stable around 3 dB so that neither spectral location as well as number of flanking bands had a influence on the amount of CMR when the masking bands were relatively wide spectrally separated.

The model simulations are shown in the right column. The simulated signal thresholds are in good agreement to the measurements. As in the measurements, a constant amount of CMR of about 3-4 dB is observed across all stimuli configurations. When flanking bands were added below the signal a two-way ANOVA (random/comodulated bands: $F(1,16) = 36.36, p < 0.01$, number of flanking bands: $F(1,16) = 5.4, p = 0.03$, interaction of the two variables: $F(1,16) = 2.21, p = 0.16$) reveals that the main effect was the envelope relation (random or comodulated). The same outcome is observed for flanking bands added above the signal (two-way ANOVA: random/comodulated bands $F(1,16) = 43.45, p < 0.01$, number of flanker bands: $F(1,16) = 0.25, p = 0.62$, interaction of the two variables: $F(1,16) = 0.44, p = 0.52$) and for symmetrically added bands (two-way ANOVA:

random/comodulated bands: $F(1,16) = 23.42, p < 0.01$, number of flanker bands: $F(1,16) = 0.83, p = 38$, interaction of the two variables: $F(1,16) = 0.45, p = 0.51$).

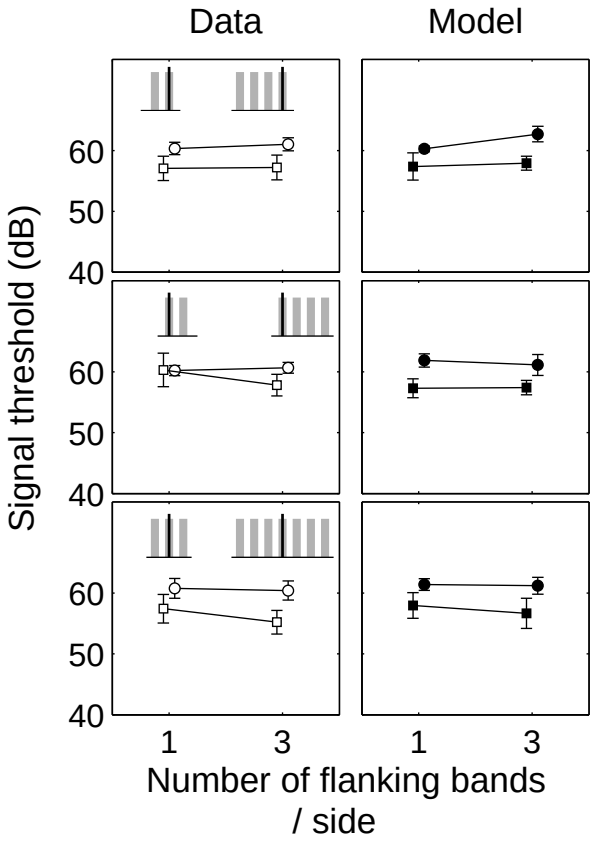

Figure 3.7: Measured and simulated signal thresholds as a function of the number and spectral location of comodulated flanking bands. The signal had a frequency of 1000 Hz. The bands had a mutual separation of one octave. Open squares denote measured, solid squares simulated thresholds. Circles show signal thresholds when noise masker were uncorrelated, squares when they were correlated. CMR is determined as the difference between uncorrelated and correlated signal thresholds. Left panels: Measured data and stimuli sketches. The gray blocks denote masking bands and the solid black line the signal. Right panels: Simulated data with the overall model.

3.9 Model analysis

Earlier studies have typically focused on specific aspects of CMR, such as characteristics of within-channel processes on CMR (e.g., Schooneveldt and Moore, 1987; Verhey et al., 1999), effects of suppression and (wideband) inhibition in CMR (e.g., Pressnitzer et al., 2001; Neuert et al., 2004; Ernst and Verhey, 2006) or across-channel mechanisms in CMR (e.g. Chaper 2 and Buus, 1985). In the present chapter, a generalized modeling framework was investigated that attempts to account for various aspects of CMR data using essentially the same model parameters across experimental conditions. In the following, the effects and interplay of the most critical processing stages and the capabilities and limitations of the modeling are discussed.

3.9.1 Effects of peripheral compression on CMR

Role of Suppression

Most of the simulations presented in this study showed a reasonable agreement with the measured data. However, clear deviations between simulations and measured data were found in Experiment 3 which considered the effect of level differences between the on-frequency band and the flanking bands on CMR. In contrast, Ernst and Verhey (2006), found a good correspondence between their simulations and the data. In order to understand the reasons for the differences between the modeling results from the two studies, additional simulations were run here using the model version "SC_NL_MFB" which considers only a single peripheral channel as in Ernst and Verhey (2006) instead of a DRNL filterbank. Also, exactly the same DRNL filter implementation as in Ernst and Verhey (2006); Plack et al. (2002) was used for peripheral filtering. This DRNL differs from the version of the present study in several parameters: The first gammatone filter in the non-linear path was tuned above the center frequency (1.1· CF) and the second filter below the center frequency (0.94· CF). Furthermore, a stronger compression was used with a compression ratio of 0.78 dB/dB for input levels below 40 dB and 0.16 dB/dB for input levels above 40 dB. These parameters were suggested in Plack et al. (2002). For comparison, the DRNL configuration used in the previous experiments of this chapter applied no compression

to input levels below 40 dB and had a weaker compression ratio of 0.25 dB/dB above 40 dB. Otherwise, the structure of the modified model remained the same as illustrated in Fig. 3.1, i.e., there was no temporal window combined with a signal-to-noise ratio based decision algorithm (as in Ernst and Verhey (2006)) but an optimal detector applied to the preprocessed stimuli.

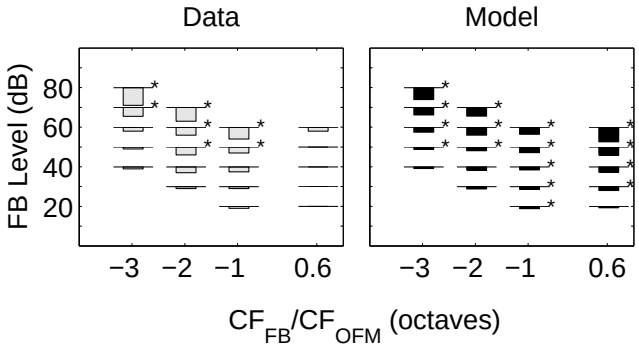

Figure 3.8: Barplot denoting the amount of CMR in a flanking band paradigm. The CMR is plotted as a function of spectral position of the flanking band relative to the signal frequency, shown on the abscissa and the level of the flanking band, shown on the ordinate. A release from masking, a positive CMR, is indicated by a bar pointing downwards. Negative CMR by a bar pointing upwards. Left panel: Measured data taken from Ernst and Verhey (2006). Right panel: simulated CMR with the auditory perception model using a compressive nonlinearity according to Plack et al. (2002).

Figure 3.8 shows the simulated results obtained with the modified DRNL stage (right panel). The left panel shows a replot of the experimental data from Ernst and Verhey (2006). Here, for flanking band frequencies below the signal frequency, the simulated data are in good agreement with the experimental data. CMR is only slightly underestimated for the highest flanking band levels at all flanking band frequencies below signal frequency. However, for flanking band frequencies above the signal frequency, CMR is overestimated. These simulations correspond closely to the model simulations in Ernst and Verhey (2006) indicating that the selection of the DRNL filter parameters critically influences the model simulations for this experiment. The differences in the detector stages are obviously less important. Since the processing in the temporal window model as used in Ernst and Verhey (2006) is more comparable

to using a modulation lowpass filter instead of a modulation filterbank a further analysis using a modulation lowpass filter instead of a modulation filterbank (model configuration: "SC_NL_LP") was run. The simulated amount of CMR did not change significantly. If however, a linear gammatone filter instead of the DRNL filter (model configuration: "SC_L_MFB") was used, no CMR at all was simulated at a any level and flanking band frequency of Experiment 3.

Taken together, the model simulations clearly demonstrate that the suppression effects caused by non-linear compression in the DRNL is the mechanism that accounts for CMR in this paradigm. The exact characteristics of the suppression/CMR effect do, however, critically depend on the specific compression coefficients assumed in the model and it is questionable whether a compression at low levels like in Plack et al. (2002); Ernst and Verhey (2006) can be justified. The compression settings of the underlying model in Jepsen et al. (2008), taken originally from Meddis et al. (2001), are not suited to account for suppression-related CMR.

Another trend that is seen in the simulations is that CMR is clearly overestimated when the flanking band is located above the on-frequency masker although suppression maps that are simulated in Plack et al. (2002, e.g.) with a single DRNL filter are consistent with experimentally obtained suppression maps in Houtgast (e.g. 1974).

The question then remains whether a non-linear filter configuration with a constant assumed compression like in the version of Plack et al. (2002) is suitable for predicting realistic suppression-related CMR at all.

Another general issue concerning suppression as simulated by the DRNL filter is that the increase in suppression with the suppressor level is directly related to the amount of compression in the on-frequency filter. However,Duifhuis (1980) showed that the amount of suppression for an increasing suppressor level is not constant but can be as high as 2.5 dB/dB. With an assumed constant compression ratio as in the DRNL filter presented here, this change cannot be accounted for. Recent modeling work (Hohmann and Kollmeier, 2007) provided suggestions to solve this problem by introducing a gain-function that depends on the instantaneous frequency and allows, e.g., the simulation of data from Duifhuis (1980). A modification of the current DRNL filter in that direction would probably also influence the simulation of suppression-based CMR.

In order to illustrate how the peripheral compression accounts for CMR in the framework of the model, the resulting suppression effect for coherent envelope fluctuations in the model's internal representation was examined. Figure 3.9 shows the differential effect of peripheral compression in the DRNL filter versus linear processing in a gammatone filter on the internal representation of the stimuli in the model. An exemplary comodulated condition from Experiment 3 was selected where the flanking band frequency was -3 octaves above signal frequency and had a level of 80 dB.

The template and reference representations are shown in the left and right upper panel, respectively. The template represents the normalized averaged difference between the internal representation of the noise plus supra-threshold signal and the noise alone serving as an internal "image" of the target signal. The reference is the averaged internal representation of the noise alone. Only the fours first modulation filters of the template and reference are shown, tuned to 0 (lowpass), 5, 10 and 17 Hz. Solid lines show the internal representations when only the on-frequency masker was presented. Dashed lines show the situation when the comodulated flanking band was added. For this analysis, 16 frozen-noise representations were averaged to obtained the template and reference. In this way, the variability of the internal representations due to random fluctuations in the stimuli was excluded and the only source of variability was left to the presence or non-presence of the flanking band.

The template shows essentially the same shape for the stimulus consisting of on-frequency and flanking band as for the on-frequency masker only. When looking at the reference representation (upper right), the modulation depth is clearly reduced when the flanking band was added to the on-frequency band (dashed lines). This effect is most salient for the low-frequency modulation frequencies mainly because the main modulation energy is distributed up to the bandwidth of the noise (20 Hz) but it can also be observed for higher modulation frequencies mainly due to the stimulus onset that excite all modulation filters.

The lower panels of Fig. 3.9 show histograms of the cross correlation coefficients between noise-alone representation and template (triangles, solid lines) and between noise-plus-signal representation and template (circles, dashed lines). In the left panel, a linear gammatone filter was used as the peripheral filter. In the right panel, the result

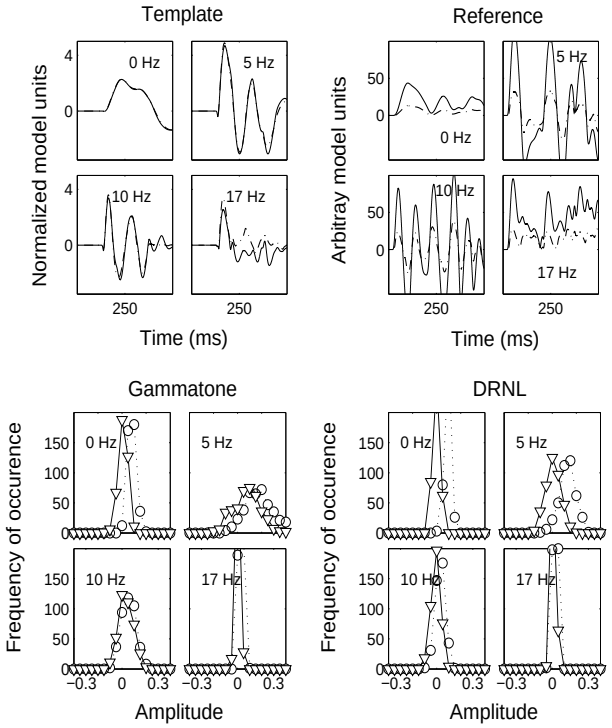

Figure 3.9: Upper panels: Internal representation at the output of the modulation filters in the on-frequency peripheral channel. Solid curves show outputs of the modulation filters presenting only the on-frequency masker. Dashed lines represents outputs when the comodulated flanking band was added. Left panels: Internal representation of the template (normalized averaged difference between supra-threshold signal and noise). Right panel: Internal representation of the average reference (noise alone without signal). 16 frozen noise representations were used for generating template respectively reference. Lower panel: Histogram of the cross - correlation coefficients at the output of four modulation filters. Correlations are calculated for exemplary condition of Experiment 3 with the comodulated flanking band located 0.6 octaves above signal frequency and with a level of 60 dB. Correlations are shown for the reference (triangles, solid line) and reference plus signal (circle, dashed line). The center frequency of the modulation filter is indicated in each panel. In terms of signal detection the reference and template are most separable the modulation filter at the output of the modulation filter at 0 Hz.

for the DRNL filter in 3.8 is shown. For linear filtering, basically no difference can be observed between the noise-alone and the noise-plus-signal distribution. The shape of the distribution changes when a non-linear DRNL filter is applied. Distributions

become narrower, particularly at the output of modulation filters tuned to 0, 5 and 10 Hz. In the case of the noise-plus-signal distribution the mean is shifted towards higher correlation values. Since the difference of their means divided by the widths of the distribution determines the separability of the condition with and without flanking band, the resulting threshold decrease in the condition with flanking band and DRNL filter is obvious from the correlation histograms. In terms of signal detection, this increase in separability is most pronounced at the output of the modulation filter at 5 and 10 Hz. The analysis revealed that first, the cue causing CMR in this type of paradigm can be regarded as a within-channel cue. Second, considering the upper panels of Fig. 3.9, this within-channel cue can be regarded as a noise reduction (sharpening of the correlation distributions) occurring mainly at the output of lower modulation frequencies of the on-frequency peripheral channel when appropriate remote masking energy suppresses on-frequency envelope fluctuations. Therefore the mechanism is termed within-channel suppression. The principle of operation described here is comparable to the effect of the EC mechanism which also can be regarded as an noise reduction mechanism (see Chapter 2).

Role of modulation cues

Beating (e.g. Chapter 2 and McFadden, 1986; Schooneveldt and Moore, 1987) caused by interacting frequency components of the stimulus that pass a single peripheral channel is considered as an important within-channel cue for CMR. In this case, adding a target signal alters the temporal envelope pattern and thus changes the characteristics of the beating which then can be exploited by the auditory system. Chapter 2 and Verhey et al. (1999) showed that CMR based on within-channel beating cues can be simulated by modulation filterbank models. Fig. 3.10 shows simulated data for the stimuli described in Experiment 4. Circles and squares represent a random and comodulated flanker, respectively. The left panel (black symbols) shows simulations from the left middle panel of Fig. 2.6 from chapter 2 where a 4th-order gammatone was used as filter. Their model thus corresponds to the model configuration "MC_L_MFB_EC" considered in the analysis here. The gray symbols represent thresholds obtained with a modified linear filter that matches the power transfer function of the DRNL filter at 65 dB SPL. Now, a asymmetry in the thresholds

for the comodulated condition is observed as well as the "plateau" in the center, as also observed in the data (from Fig. 3.5). The simulations with the single "broader" filter suggest that within-channel cues appear to be salient up to a relative frequency separation of 0.7. Since no explicit across-channel EC mechanism was applied here, no CMR can be simulated at the lowest flanking band frequency (corresponding to a frequency ratio of 0.5). The black symbols in the right panel of Fig. 3.10 are a replot of the DRNL simulation from Fig. 3.5. It was found in a statistical analysis that the main detection cues is based on temporal beating cues as in the statistical analysis similar to those in Fig. 2.7 from Chapter 2 using a gammatone filter. If the beating information is removed from the analysis by replacing the modulation filterbank by a modulation lowpass filter, the amount of simulated CMR is strongly reduced, as indicated by the gray curves and gray symbols in the same panel. Still, there is a remaining small effect for some of the spectral separations when no modulation (beating) cues are available. The remaining CMR can be attributed to suppression caused by peripheral compression followed by temporal averaging. A comparable processing has been suggested by Buschermöhle et al. (2007) to account for within-channel CMR without the need for modulation filters. In their model, the mean envelope was compressed by a power-law function constant exponential factor. The compressed envelope was then averaged over time which corresponded to a modulation lowpass-filter processing. The modified model used here combining the non-linear compression of the DRNL filter and a modulation lowpass filter, behaves similar to the analytic model of Buschermöhle et al. (2007).

The analysis indicates that, in the framework of the present model, instantaneous compression does not change the main detection cue which is based on temporal information exploited by the modulation filterbank. Therefore, compression alone cannot alone account for the CMR in this type of paradigm which is consistent with the results from the modeling study of (Buschermöhle et al., 2007). However, the observed asymmetry of signal thresholds is an effect of the increased filter bandwidth (relative to that of the gammatone filter) which in turn is a consequence of level dependent peripheral compression and is accounted for by the DRNL filter. Thus, the use of a non-linear and level dependent auditory filter like the DRNL filter enlarges the influence of within-channel cues in the model for some experimental conditions.

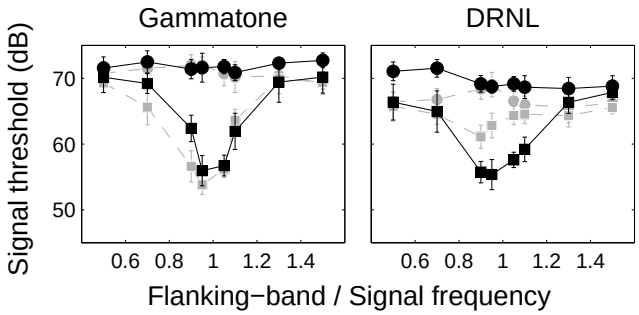

Figure 3.10: Simulated signal thresholds for two different model configurations. Circles show signal thresholds when masker bands were uncorrelated, squares when bands were correlated. Left panel: Solid symbols show simulated thresholds replotted from Fig. 2.6 in Chapter 2 performed with a model configuration of "MC_L_MFB_EC". Gray symbols show simulated thresholds when a combination of a single gammatone and lowpass filter was used as filter. The combined filter had a magnitude transfer function compared to a DRNL filter at around 65 dB SPL. Right panel: Solid symbols indicate simulated data replotted from Fig. 3.5 obtained by a "MC_NL_MFB" model configuration. Gray symbols show signal thresholds when a lowpass filter instead of a modulation filterband was applied similar to Buschermöhle et al. (2007). Increasing CMR with decreasing mutual spectral distance from bands can be observed with all configurations.

Interaction of compression and EC process

Fig. 3.11 shows the effect of the EC process on the simulated thresholds in Experiment 2. The black symbols are replotted from Fig. 3.3 (left panel). Additionally, the gray symbols are simulated signal thresholds when the EC mechanism was switched off. Without the explicit across-channel mechanism, no CMR was simulated in any of the level conditions. In Chapter 2 it was demonstrated, using the EC model version with the gammatone filterbank, that the decision advantage in the model due to the EC mechanism is reflected by a reduction of the noise floor at the output of the various modulation filters (see Fig. 2.4 in Chapter 2). A similar analysis with the current model using a DRNL filterbank revealed that such behavior is not affected largely by the level dependent peripheral stage. Thus, the compressive nonlinearity does not affect the simulated CMR associated with the EC processing in the conditions considered in the present chapter.

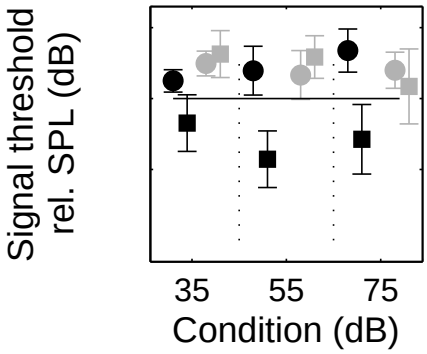

Figure 3.11: Circles illustrate signal thresholds when the masker bands were uncorrelated, squares when they were correlated. The solid circles show simulated signal thresholds replotted from Fig. 3.4 when the EC mechanism was turned on. The gray circles show signal thresholds in the case when no EC mechanism was applied. No CMR is simulated in that case.

3.9.2 Multi-channel vs. single-channel processing

In the present chapter (Chapter 3), model simulations using all auditory filters without a-priori knowledge of the stimuli and thus without telling the model "where to listen". In contrast, model simulations in conditions with narrowband target signals (e.g. a tone) and earlier versions of the model often used a few auditory filters spread narrowly around the target frequency (e.g. Dau et al., 1997a; Jepsen et al., 2008). A few filters were used to still account for off-frequency listening while filters that presumably do not contain valuable information about the target signal were excluded from the model's processing and decision stage. Here, the usage of all auditory filters led to some discrepancies of the simulations and the data in Experiment 1 (broadband masker). In contrast, earlier "single channel" simulations (e.g. Chapter 2 Verhey et al., 1999) of the same condition using a-priori knowledge to select the target-signal centered band were successful in predicting the data. In order to clarify the discrepancy between measurements and simulations of the present model in Experiment 1, additional simulations were performed with different model configurations as shown in Fig. 3.12. The upper rows shows simulated thresholds

when a single filter (model configuration "SC_L_MFB") was used. The lower rows show simulations using a filterbank (model configuration: "MC_L_MFB"). The single filter was always centered on the target signal frequency of 2 kHz. The configuration with a single gammatone filter centered to the signal frequency of 2000 Hz (upper left panel) predicts the data sufficiently well (see Fig. 3.2). Signal thresholds drop from -10 dB to 10 masker spectrum level and remain constant at around 15 dB for the other spectrum levels similar to the measured data.

When a single DRNL filter was used for detection (upper right panel) three trends can be observed. First, signal thresholds in the uncorrelated case rise from around 21 dB at -10 dB masker spectrum level to 25 dB at 40 dB spectrum level. This is related to the increasing filter bandwidth for higher levels (Jepsen et al., 2008) resulting in more masker energy that falls into the passband of the filter. Second, for the lowest spectrum level of -10 dB, a threshold of around 17 dB is observed which is significantly lower than in the case of the gammtone filter. In this -10 dB level condition and the gammatone filter model, the thresholds in both the comodulated and random case were limited by the absolute threshold realized in the model. The expansion stage within the DRNL filter increases the distance of the target level and the threshold (doubles the target level on a logarithmic scale)and thus abolishes the limitations by the absolute threshold. A repeatition of the simulations at -20 dB spectrum level (not shown) showed again comodulated thresholds at about 21 dB. Third, comodulated thresholds drop from 17 dB to 15 dB for the lowest spectrum levels and then increase to 19 db respectively 21 dB for the highest spectrum levels. For the highest two spectrum levels, uncorrelated and comodulated thresholds show a parallel tendency to rise. The rising thresholds coincide with the transition between the linear and compressive part of the broken-stick nonlinearity is reached between 10 and 30 dB spectrum level. When the compression is starting to act and the dynamic range of thresholds is decreased.

In the multi-channel configuration (lower row) the model fails to account for the observed CMR with both the gammatone and the DRNL filter. No clear separation can be observed between the signal thresholds for the random and for the comodulated condition. This suggests that mainly the transition from a single-channel to a multi-channel configuration is deteriorating the salience the within-channel cues used in the

single-channel model and therefore the ability of the multi-channel model to predict CMR. Furthermore this deterioration is only observed when a broadband noise is used as masker (see also Experiment 3.6). The reason can be found in the way the template is generated and how the decision device combines information across filters. The template is averaged over a finite number (16) of different supra-threshold signal representations. For non-deterministic stimuli, it resembles therefore a "noisy" estimate of the true target signal representation, with the degree of "noisyness" depending on the number of averages. The multi-dimensional template is then cross-correlated with the actual noise-plus-signal representation and with the noise-alone representation, respectively. Both resulting correlation values can be interpreted as a detectability index that equals the square root of the sum of squares of the detectability indices in the individual auditory filters (Dau, 1996), hence optimally combining the information across filters. Multi-channel simulation were proven to indeed act as an optimal detector in modulation detection experiments (e.g. Dau et al., 1997b), where across-channel information of modulation patterns was combined. In such a situation valuable information about the target (modulation) was available in all auditory channels and the template estimated the true target sufficiently well with a limited number of averages. In the current experiment, where a tone is to be detected in a broadband noise (random or comodulated), the template mainly consists of valuable target activity in a single (on-frequency) auditory channel, while noise is dominating the other channels. In a single-channel simulation, the "correct" channel is selected using a-priori knowledge and the data can be predicted. However, if a multi-channel simulation is performed, all channels excluding the on-frequency channel in the template mainly carry a residual, non-zero estimate of the noise. Since this residual "noise" in the template causes a non-zero correlation with external noise in the stimuli, the detection performance of the model diminishes. Hence thresholds in the comodulated condition approach thresholds for the random condition. Since the fingerprint of the broadband noise is still left in the off-frequency filters of the template acts as an additional "external noise" parameter that limits performance analogue to the internal noise. The gray symbols in the lower panels of Fig. 3.12 show simulated detection thresholds in the case when 16000 instead of 16 repetitions were performed for template generation. In this way the residual noise contained in the template

was stronlgy reduced. Simulations are shown only for a spectrum level of 10 dB where the largest deviation from the single channel model was observed. It can be seen that simulated threshold approach those obtained when only a single gammatone filter is used which is the best approximation to the measured data. By averaging over this large number of repetitions the residual noise in the template is decreased (see e.g. Snedecor and Cochran, 1989) by a factor of $10\,log(\frac{16000}{16}) = 30$ dB and cross-correlations of the template with the noise-plus-signal respectively noise-alone representation in the off-frequency filter channels approach zero and do not influence the simulated data. It can be concluded that the discrepancy between experimental data and simulations in Experiment 1 are solely due to the nature of the decision device and in principle can be accounted for by the present pre-processing stages in the model.

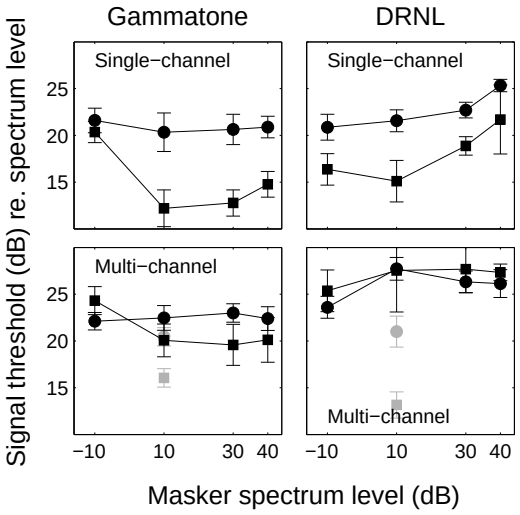

Figure 3.12: Signal thresholds as a function of the masker spectrum level when the masker band was uncorrelated (circles) and when it was correlated (squares). Left column panels: Simulated data when a gammatone filter was used as a peripheral filtering stage. Right column panels: Simulated data when a DRNL filter was used as a peripheral filtering stage. Upper row panels: Simulated data when a single filter was used for detection. Lower row panels: Simulated data when a multichannel configuration was used for detection. The gray symbols in the lower panels show simulated data for a multichannel configuration when the template was averaged 16000 instead of just 16 times.

3.9.3 Characteristics of the across-channel processing

In a flanking band type of paradigm the amount of CMR depends on the spectral separation of the flanking bands. For "relatively" large spectral separations, depending on the frequency and (relative) levels of the bands, only across-channel processing contributes to CMR. For smaller separations CMR is mainly based on temporal cues (e.g. Schooneveldt and Moore, 1987; Hall et al., 1990). In Chapter 2, a hypothetical EC process was introduced and assumed to be level independent. In addition, it was assumed that the CMR resulting from the EC process is also independent of the number of flanking bands as long as pure across-channel processes are involved. This is due to the equalization stage of the EC process where the energy falling in each filter is summed and divided by the total energy of the components. The results of the experiments and simulations of the present study support the validity of this concept. In experiment 5 (Fig. 3.7) the same amount of CMR of about 3-4 dB was found for 1,3 or 6 flanking bands. Thus, in this framework, CMR effects larger than about $3-4$ dB would imply that additional within-channel processes based on beating cues or suppression are involved.

3.10 Summary and Discussion

3.10.1 Experimental findings

A number of novel experimental findings have been obtained in this study. In Experiment 1 CMR was determined in a broadband condition of the bandwindening experiment as a function of the masker spectrum level. It was found that apart from relative low spectrum levels when the absolute threshold was a limiting factor, signal thresholds re masker spectrum level were independent of the masker spectrum level. These findings are consistent with Moore and Shailer (1991); Hicks and Bacon (1995); Hatch et al. (1995). In the study of Hatch et al. (1995), CMR was measured in a comparable paradigm where 20-Hz wide bands were separated by 400 Hz. Three, five, or nine flanking bands were presented symmetrically around the signal frequency of 2000 Hz. A constant CMR of around 5 dB was measured across all masker

configurations which is in good correspondence good to the (slightly larger) amount of CMR found here although the amount of CMR is slightly larger.

Secondly, in Experiment 2 CMR was measured in a flanking band type of paradigm for one-octave separated narrow masker bands. The overall level of the bands was either 35, 55 or 75 dB SPL. The amount of CMR remained constant around 4 dB suggesting that the cue in this type of paradigm is not susceptible to overall level.

At last, in one part of Experiment 5 CMR was measured as a function of the number of flanking bands when the masking bands had a one-octave spacing. The wide spacing was selected to ensure that mostly "'true"' across-channel cues were responsible for the observed data. It was found that a rather constant amount of CMR of around 3-4 CMR was measured in these conditions, independent of the number of flanking bands. The results are in line with findings from Chapter 2 and Schooneveldt and Moore (1987). It also supports the assumptions that the "'true"' across channel component of CMR is quite constant and not sensible to parameters like number of masking components or level (see above).

3.10.2 Modeling

In the previous sections a computational auditory processing model which predicts thresholds in various paradigms that give rise to CMR was described. The comparison of data and model predictions suggest that at least three mechanisms contribute to the overall CMR: within-channel beating, within-channel suppression and an across-channel EC-type mechanism.

The model structure proposed here is based on the model from Dau et al. (Chapter 2 and 1997a) which incorporates mechanisms to account for within-channel beating cues and the across-channel processing. The model, so far, could not account for within-channel suppression and the resulting CMR effects. In order to extend the existing model of see Chapter 2 to also account for within-channel suppression, the linear gammatone filters that represents the cochlear processing were replaced by DRNL filters as already integrated in the companion model of (Jepsen et al., 2008). The so extended model is able to also show CMR in experimental conditions where suppression was suggested as the main underlying mechanism. However, there were more deviations between data and model predictions as in the study by Ernst and

Verhey (2006). One important difference between the approach in Ernst and Verhey (2006) and the resulting DRNL-based model here is the decision device. Ernst and Verhey (2006) used a sliding temporal window where signal thresholds corresponded to a fixed signal-to-noise ratio at the output of the temporal window. The signal-to-noise ratio in a run was determined as the ratio of the maximum in the interval containing the masker-plus-signal and the maximum of the intervals containing only the masker. This detector combines single instants in time at different temporal positions of the observation interval to derive a threshold. However, it is unclear how the comparison of non-simultaneous "local features" in time can effectively take place without assuming a memory. Moore et al. (1988) for example applied the largest *simultaneous* signal-to-noise ratio as the decision criterion. In contrast, the present model uses an optimal detector (Dau et al., 1996; Green and Swets, 1966) as the decision device. Memory is provided in form of a template which forms a fingerprint of the target signal features to listen for. The template is then correlated with the actual representation of the stimulus introducing a possibility to account for temporal integration of information which is comparable to the "multiple-look" theory from Viemeister and Wakefield (1991).

Although the detectors applied in both approaches are quite different, the main reason for the deviations of the model predictions between both studies appears to be in the specific choice of parameters in the DRNL stage as analyzed in Section 3.9.1. It was shown that suppression-based CMR is very sensitive to the compression ratios applied within the DRNL filter. For a release from masking it is essential to have a compressive I/O function for input levels up to 40 dB (transition point), which is the case in Ernst and Verhey (2006) and the present study. In the present model, however, no CMR is predicted below the transition point were no compression is applied using the same DRNL configuration as in Jepsen et al. (2008). In the model analysis in SecMMM and in Ernst and Verhey (2006) compression was also applied for low levels based on findings of Oxenham and Plack (1997). They measured basilar membrane compression using a forward masking paradigm and estimated compression ratios of 0.78 dB/dB for input levels lower than 40 dB and 0.16 dB/dB for input levels above. Physiological studies (e.g. Robles and Ruggero, 2001) suggests that responses to CF tones grow linearly at low stimulus intensities which is in contradictions to the above

mentioned psychoacoustical findings. Taken together, suppression based CMR can not be fully accounted for by the DRNL parameters which were successfully used to account for a large set of other experimental data in the companion model of Jepsen et al. (2008).

A major advantage of the model suggested here is the generality of the approach. The use of the optimal detector enables the model to use any experimental input stimulus that fulfills a signal-plus-noise configuration in a wider sense. Since the major processing parts are shared with the successfully proven model of Jepsen et al. (2008), the model is compatible with various other monaural experimental conditions. Similarity of the model framework exist also to binaural models of Breebaart et al. (2001a); Thompson and Dau (2008), while the EC-process used to account for across-channel CMR resembles some basic "classically" binaural model approach. The inclusion of a non-linear peripheral filterbank offers future perspectives to study the effect of hearing loss in CMR conditions and to potentially separate effects directly related to peripheral deficits and more central ones in complex listening situations.

Limitations of the current model approach might occur due the choice of the DRNL filter which inherently links suppression and compression. The compression above the transition point within a DRNL filter remains constant (Jepsen et al., 2008) and is not able to show suppression more than -1 dB per dB suppressor level (Hohmann and Kollmeier, 2007). Additionally the amount of suppression would remain constant which contradicts findings from Duifhuis (e.g. 1980). In his study data show that the change of suppression is not constant with increasing suppressor level. A possible solution was given in Hohmann and Kollmeier (2007) by introducing an instantaneous frequency dependent gain function.

Another limitation is the currently "hard-wired" EC process to account for across-channel CMR. While the two identified within-channel mechanisms linked to beating cues and compression-based suppression can be assumed to be indeed hard wired, the remaining across-channel process appears to be modulated by auditory grouping/streaming constraints (Dau et al., 2009). The next chapter is partly concerned with this phenomenon.

4

Effects of auditory grouping on modulation perception

4.1 Abstract

Coherent modulations in one sound can aid in the detection of another sound, as was demonstrated in conditions of comodulation masking release (Hall et al., 1984) and binaural masking release (Durlach, 1963). On the other hand, modulation in one frequency region can also impede the detection or discrimination of modulation in other frequency regions (Yost et al., 1989). Although the neural substrates for across-frequency modulation processing remain unclear, recent studies have concentrated on brainstem structures (Pressnitzer et al., 2001). In this study, it is shown that sounds occurring after the target sound in time determine whether or not across-frequency modulation effects are observed. The results suggest that the binding of sound elements into coherent auditory objects precedes aspects of modulation analysis and imply a cortical locus involving integration times of several hundred milliseconds. In other words, the modulation analysis necessary for signal detection is performed on objects, rather than frequency channels. The results place strong constraints on the search for neural correlates of this important aspect of auditory processing, and on future models of spectrotemporal processing.

4.2 Introduction

Across-frequency comparisons of temporal envelopes are likely to be a general feature of auditory pattern analysis and may play an important role in extracting signals from noisy backgrounds, or in separating competing sources of sound. Comodulation of

different frequency bands in background noise facilitates the detection of tones in noise, a phenomenon known as comodulation masking release (CMR). CMR has been measured in two ways. The first is to use a single band of noise, centered around the signal frequency, as a masker and to compare thresholds for modulated and unmodulated maskers as a function of the masker bandwidth (Hall et al., 1984; Haggard et al., 1990; Schooneveldt and Moore, 1989; Carlyon and Stubbs, 1989). The second method is to use a masker consisting of several narrow masker bands, one at the signal frequency (on-frequency band) and one or more flanking bands spectrally separated from the on-frequency band (Hall et al., 1984; Schooneveldt and Moore, 1987).

CMR is usually assumed to depend on comparisons of the outputs of different auditory filters. However, especially in conditions using a single band of noise, within-channel cues can also facilitate the detection of a signal in modulated noise, leading to a CMR without any necessary involvement of across-channel mechanisms (Schooneveldt and Moore, 1987; Verhey et al., 1999). In order to establish a physiological basis for CMR, it is important to separate within-channel from across-channel effects. Similarly, it is important to determine the extent to which CMR is affected by auditory grouping processes, which are unlikely to be peripheral in nature. For instance, Pressnitzer et al. (2001) and Neuert et al. (2004) recently presented physiological data at the level of the cochlear nucleus (CN) in support of a possible physiological implementation for a model of CMR at a relatively peripheral processing stage. In contrast, at least two psychophysical studies have suggested that CMR may interact with auditory object formation, commonly associated with higher-level processes (Grose and Hall, 1993; Dau et al., 2009). The first part of the present study investigates the influence of concurrent and sequential streaming cues on CMR. The hypothesis was that perceptual segregation or grouping may affect across-channel CMR while it should not affect within-channel CMR. If so, the differential effect of grouping cues may provide a functional definition of within- and across-channel CMR.

The second part of the study examines effects of sequential streaming on binaural masking release. It has been suggested that the same detection cue (envelope cross-correlation) might be involved in monaural, across-channel CMR and in envelope-

based, binaural masking level differences (BMLD). If CMR depends on auditory grouping, the question is whether this could be a common principle for across-channel envelope processing, in terms of monaural across-auditory channel and binaural across-ear processing. Here, effects of sound occurring after the target sound in time are investigated in particular.

The third part of the study investigates an example where across-channel processing leads to a deterioration in performance in contrast to the increase in performance observed in CMR and BMLD. In modulation detection interference (MDI), the ability to detect amplitude modulation in one frequency region (target) can interfere with the presence of amplitude modulation in another frequency region (masker). The question is if perceptual segregation of the probe and the masker can in this case lead to enhanced detection by removing the detrimental effect of the modulated maskers. As in BMLD, the particular emphasis is put on the effect of sound occurring after the target sound. The results are finally discussed in terms of principal processing strategies underlying modulation perception, possible neural substrates of across-channel modulation processing, and consequences for future models of auditory modulation perception. Possible consequences for speech perception in complex environments in hearing-impaired listeners are also discussed.

4.3 Experiment 1: Comodulation masking release

4.3.1 Rationale

In this part, the effect of auditory grouping on comodulation masking release was investigated. In a flanking-band paradigm, four postcursors were presented after the flanking bands in order to form a sequential stream with the flanking bands. If across channel processing takes place at more central stages, CMR ahould be dimished when postcursors are presented. In order to investigate the possibility that the effect of the postcursors is due to attentional confusion, a control experiment was run in which the postcursors are shifted by a half an octave from the frequencies of the flaning bands. If the effect of the postcursors previously was due to a confusion effect, then CMR should be reduced also; however, if the effect was due to auditory grouping of the

flanking maskers and postcursors into one single stream, then the amount of CMR should be the same as completly without any postcursors, because the postcursors should no longer form a stream with the flanking masker bands.

4.3.2 Method

Apparatus and procedure

An adaptive, three-interval, three-alternative forced-choice (AFC) procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point of the psychometric function. The initial step size was 8 dB, which was reduced to 4 and 2 dB after the second and fourth reversals, respectively. Threshold was defined as the mean of the levels at the last six reversals of a threshold run. Four threshold estimates were obtained from each listener in each condition, final thresholds were calculated as the mean of these four estimates per subject.

Subjects

Six normal-hearing listeners ranging in age from 25-39 years participated. All had prior experience in psychoacoustical experiments.

Stimuli

The signal was a 1000-Hz pure tone, 187.5 ms in duration including 20-ms raised-cosine ramps. The composite noise masker consisted of five bands of noise, each 20 Hz wide. In the narrowband configuration, the noise bands were centered at 794, 891, 1000, 1123 and 1260 Hz representing a sixth-octave spacing around the signal frequency. In the broadband condition, the noise bands were centered at 250, 500, 1000, 2000 and 4000 Hz, covering a frequency range of 4 octaves with a one-octave spacing between the bands. In both configurations, the noise bands were generated in the time domain and restricted to the appropriate bandwidth in the Fourier domain. Comodulated noises were frequency-shifted versions of the on-frequency band. The level of each of the noise bands was 60 dB SPL. Thresholds for the 1000-Hz tone masked by the noise band centered at 1000 Hz were measured in eight conditions,

which are illustrated in Fig. 4.1: (1) Random flanking bands (R); (2) Comodulated flanking bands (C); (3) Random flankers followed by four "postcursors" (PR), or following bands; (4) Comodulated flankers followed by four postcursors (PC).

When broadband flanker were used four additional conditions were used as a control experiment as seen in (5) Random flanking bands followed by postcursors where the first postcursor was removed (GPR); (6) Random flanking bands where the center frequencies of the postcursor bands were shifted relative to the flanking band frequencies (OPR); (7) Comodulated flanking bands followed by postcursors where the first postcursor was removed (GPC); (8) Comodulated flanking bands where the center frequencies of the postcursor bands were shifted relative to the flanking band frequencies (OPC); The off-frequency post-cursors had a half-octave separation from the respective flanking bands. i.e. at 354, 707, 1414 and 2828 Hz. The four pre- or postcursors were all the same duration as the on-frequency band and were separated by gaps of 62.5 ms, giving an overall repetition rate of 250 ms, except in condition (5) when the first post-cursor was removed.

4.3.3 Results

Panel (a) of Fig. 4.2 shows the mean data across subjects for the first four conditions (R, C, PR, PC). The open squares indicate thresholds obtained for the narrowband configuration and the open circles represent thresholds for the broadband configuration.

Panel (b) of Fig. 4.2 shows the threshold release obtained with comodulated versus random flanking bands. These differences represent the amount of CMR. In the narrowband configuration, thresholds are always lower in the comodulated condition than in the corresponding random condition (left panel), leading to a CMR (open bar plots right panel) of 9.4 dB (paired t-test; $t(5) = 8.46$, $p < 0.001$) when the threshold in the random condition was compared to those in the comodulated condition (R-C), 8.2 dB when thresholds of the random condition and the comodulated flanker were followed by postcursors (R-PC). When compared to the random condition with postcursors, (PR-PC) the amount of CMR was 8.0 dB. No significant reduction in CMR was produced when the postcursors were added (condition C-PC, paired t-test: $t(5) = -1.06$, $p = 0.34$). Overall, a highly significant CMR effect was observed

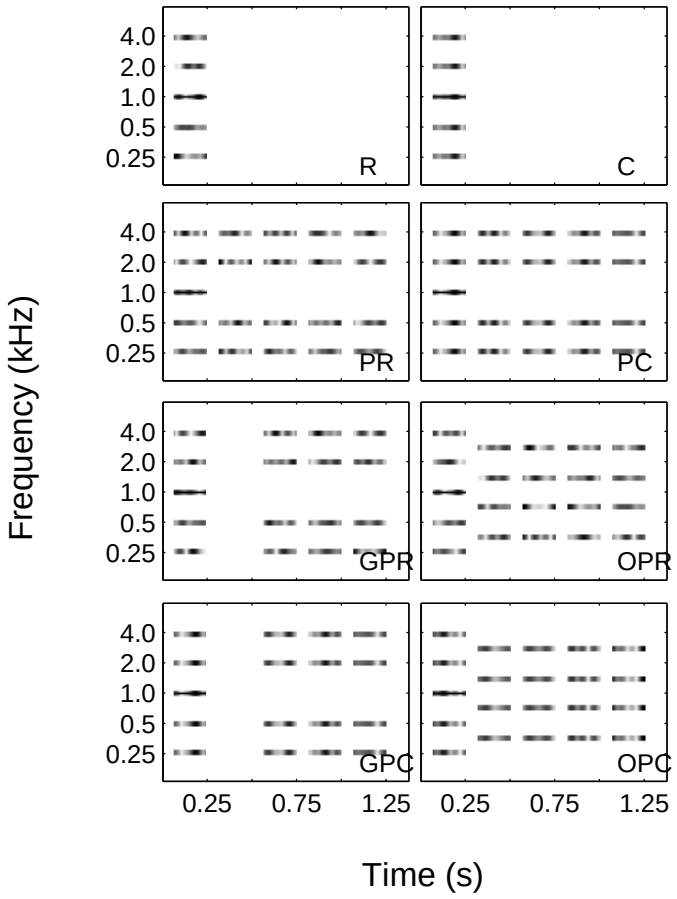

Figure 4.1: Sketches of the experimental conditions. Eight conditions with a 1-kHz target tone (black horizontal line) masked by a noise band centered at 1-kHz with random flankers (R), comodulated flankers (C), random flankers followed by four postcursors (PR), comodulated flankers followed by four comodulated postcursors (PC), random flankers followed by four poscursors where the first postcursor was left out (GPR), comodulated flankers followed by four comodulated postcursors where the first postcursor was left out (GPC), random flankers followed by four postcursors with shifted center frequencies (OPR), comodulated flankers followed by four postcusors with shifted center frequencies (OPC). The shades of gray indicate the distribution of envelope fluctuations in the masker and flanker bands.

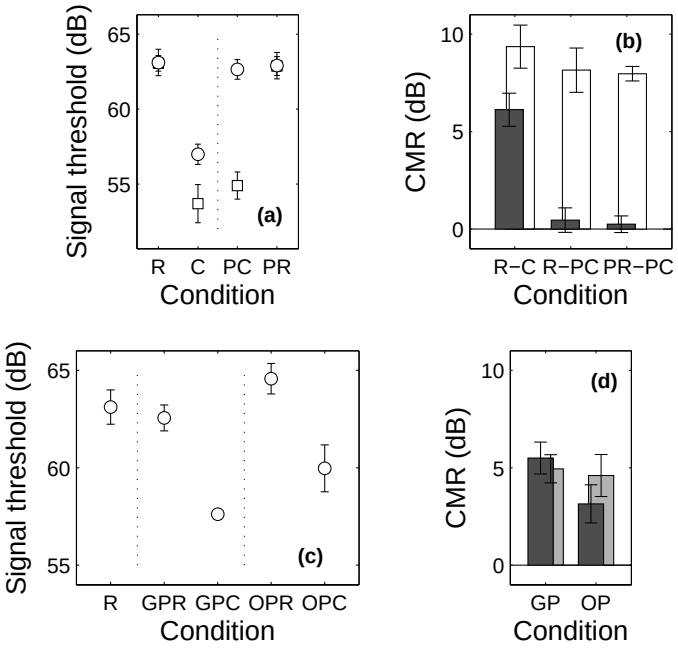

Figure 4.2: Panel (a) shows mean masked thresholds four the conditions (R, C, PR, PC) for the target tone for the broadband (circles) and the narrowband configurations (squares). Error bars denote one standard error across subjects. Conditions are indicated on the abscissa, R=random modulations of the flanking bands, C=comodulated flanking bands, PC =postcursors with comodulated flanking bands, and PR=postcursors with randomly modulated flanking bands. Panel (b) shows the amount of CMR, defined as the difference between thresholds in the random and the comodulated conditions, is indicated for the standard condition without postcursors, R-C, and for the conditions with postcursors (R-PC) and (PRŮPC). Panel (c) shows mean masked thresholds for the four control conditions when the masker bands were broadly separated, GPR=random flankers with first postcursor omitted, OPR=random flankers with frequency shifted postcursors, GPC=comodulated flankers with first postcursor omitted, OPC=comodulated flankers with frequency shifted postcusors. Panel (d) shows the corresponding amounts of CMR for the conditions GP and OP. Dark-gray bars indicate the conditions R-GPC and R-OPC and light-gray bars indicate the conditions GPR-GPC and OPR-OPC.

in the narrowband configuration, independent of condition. This is indicating that postcursors did not significantly affect target detection in the narrowband configuration, where CMR is more likely to be based on withinchannel cues.

The results are different for the broadband configuration (panel (a) open squares).

While there was a clear CMR effect (filled bar plots right panel) of about 7 dB in the standard condition (R-C), no CMR was observed in any of the other conditions. Thus, as can been seen by the bar plot in the right panel of Fig. 4.2, CMR was essentially eliminated in conditions promoting perceptual segregation of the on-frequency and flanker bands.

The results from the broadband configuration show a very strong influence of factors designed to affect the perceptual grouping of the masker and flankers. All attempts to perceptually segregate the on-frequency masker from the flankers resulted in a complete elimination of CMR, suggesting that "true" CMR as believed to be observed in the broadband configuration, does not act in isolation from the processes giving rise to auditory object formation. This is probably not compatible with peripherally based explanations of CMR in terms of, for instance, processing in the cochlear nucleus (e.g. Pressnitzer et al., 2001). Especially difficult to explain in peripheral terms is the strong influence of the temporally following bands, or "postcursors," which seem to influence the perception of the masker and flankers after their presentation. This effect is unlikely to be related to backward masking, as the time scale was much greater in this case; backward masking is generally negligible within about 20 ms (e.g. Oxenham and Moore, 1994). In contrast, the results from the narrowband configuration showed no effect of grouping manipulation.

To further stress the possibility that the effect of the postcursors in the broadband condition was indeed due to perceptual auditory grouping, the first burst of postcursors following the flankers was replaced with a gap of silence (condition GPR and GPC Fig. 4.1). The data in panel (c) in Fig. 4.2 show that CMR is not affected by the postcursors, if the first postcursor is omitted. The threshold obtained in the comodulated condition with gap-postcursor (GPC) was not significantly different from the threshold obtained in the original condition C (from panel (a)). The corresponding CMR is shown in panel (d). CMR was 5.5 dB when defined as R-GPC (dark-gray bar), and 4.9 dB when defined as GPR-GPC (light-gray bar). The results from the second control experiment, where the postcursors were presented at intermediate frequencies off-frequency postcursors show slightly elevated thresholds both for the random (OPR) and the comodulated (OPC) conditions compared to the standard thresholds, R and C. However, the amount of CMR, as measured by the difference

between OPR and OPC conditions (right-hand light-gray bar); 4.6 dB, was within 1.5 dB to that found in the standard condition without postcursors, R-C.

Taken all this observations together, these results support the hypothesis that two different mechanisms contribute to CMR: results from the broadband conditions reflect across-channel mechanisms, which are susceptible to grouping manipulations, while the results from the narrowband conditions reflect within-channel mechanisms, which do not depend on variations of the acoustical context. Thus, the introduction of stimulus components and sequences as described here allow one to separate within- and across-channel contributions to CMR.

4.4 Experiment 2: Binaural masking level differences

4.4.1 Rationale

Several phenomena have been investigated that facilitate the perception of masked sound as there are binaural masking level differences (BMLD) and monaural comodulation masking release (CMR). In Experiment 1 it was shown that across-channel CMR is susceptible to perceptual auditory grouping. Since it is assumed that both across-channel CMR and high-frequency BMLD use envelope correlation between different channels as a detection cue (e.g. van de Par, 1998) high-frequency BMLD should also be susceptible to perceptual grouping. In this part BMLD was measured with three different stimuli configurations: (i) high-frequency BMLD with and without postcursors (see Experiment 1); (ii) low-frequency BMLD with and without postcursors and (iii) BMLD with transposed stimuli. Transposed stimuli were introduced by van de Par and Kohlrausch (1997) in order to preserve "'fine-structure"' information that is normally lost by basilar-membrane filtering and hair-cell transduction at high frequencies. Thus the temporal low-frequency information is now available at high-frequencies and the influence of "'fine-structure"' processing on auditory grouping can be evaluated.

4.4.2 Method

Apparatus and procedure

The same adaptive, three-interval, three-alternative forced-choice procedure as in experiment 1 was used.

Subjects

Six normal-hearing listeners ranging in age from 25-39 years participated in the configuration when BMLD was measured for high-frequency and low-frequency masking bands. Four of this subjects participated in the case when transposed stimuli were used. All subjects had prior experience in psychoacoutical experiments.

Stimuli

BMLD was measured at a signal frequency of 5000 Hz in the high-frequency conditions and at a signal frequency of 500 Hz for the low-frequency conditions. As in the CMR conditions, the Gaussian noise masker had a bandwidth of 20 Hz, a duration of 187.5 ms including 20-ms ramps, and a level of 60 dB SPL. Four stimulus conditions were investigated, as indicated in Fig. 4.3.

. In the single-band (reference) condition, signal and masker were presented only to the left ear. In the standard condition, the signal was presented in anti-phase to the left and right ear and added to the identical noise masker band in both ears (N_0S_π) in order to create highly similar envelopes. In the postcursor condition, four postcursors were presented after the masker in one of the two ears whereby stimulus and gap durations were the same as in the corresponding CMR condition. Finally, an "off-frequency" postcursor condition was tested in a control experiment, where the postcursors were presented at 2815 Hz (high-frequency condition) and 281 Hz (low-frequency condition), respectively. For the transposed stimuli a 25-wide Hz wide gaussian noise with a center frequency of 250 Hz was generated. When needed the signal was added in anti-phase to the noise to create a N_0S_π listening condition comparable to that before. The stimuli then was half-wave rectified and lowpass-filtered with a cutoff frequency of 1500 Hz in order to mimic haircell lowpass

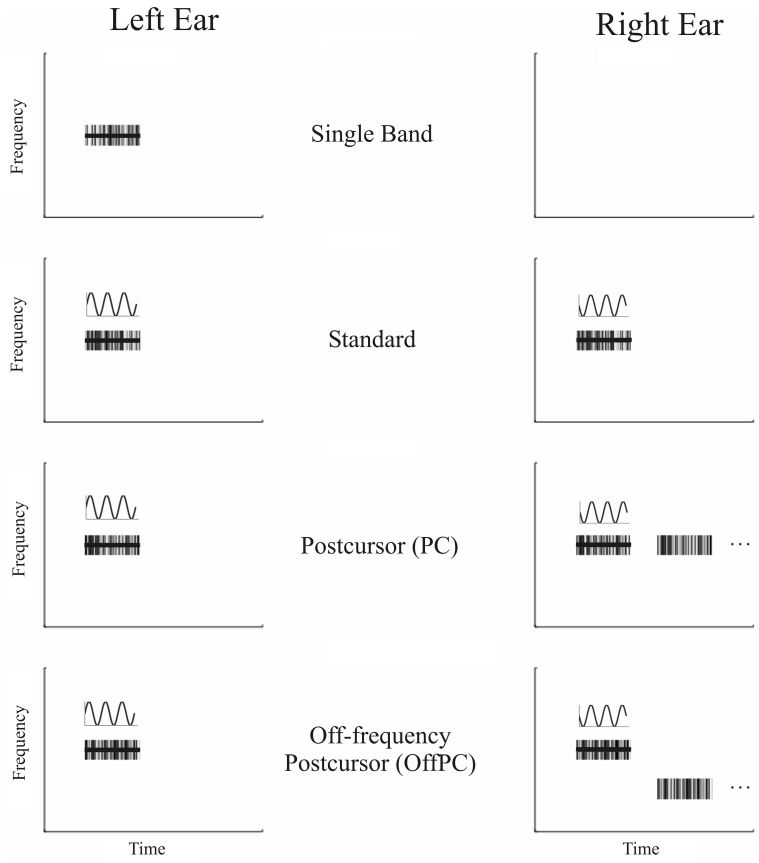

Figure 4.3: Four stimuli configurations that are used to investigate the influence of auditory grouping on high-frequency BMLD. The patterned boxes represent the masker whereas the test signal is shown as the solid line within the masker. The waveform above the masker determines the phase of the sinusoid. The signals on the both ears have a phase shift of 180 degree. The single band condition serves as the reference from which the BMLD are calculated.

behavior (Palmer and Russell, 1986). The so obtained "'pre-processed"' stimuli is finally multiplied with a 5000 Hz tone. Postcursors were generated in the same way but for each bands an independent noise source was used.

4.4.3 Results

The results at high frequencies are shown in Fig. 4.4. The left panel shows BMLDs for the individual subjects obtained in the three conditions (standard (STA), on-frequency postcursors (PC), off-frequency postcursors (OffPC)).

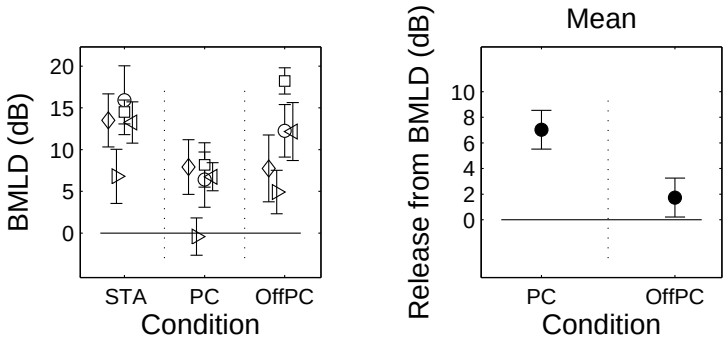

Figure 4.4: The left panel shows the individual binaural masking level differences (BMLDs) for five subjects, obtained at 5000 Hz signal frequency in three conditions (standard "STA", postcursors "PC", and off-frequency 2800-Hz postcursors "OffPC"). The right panel shows the average release from BMLD by auditory stream segregation in the two conditions "PC" and "OffPC".

There is a considerable variability across subjects in the amount of BMLD, consistent with other data from the literature (e.g. van de Par and Kohlrausch, 1998) e.g. one subject (right handed triangles) shows lower overall BMLD but a similar release from BMLD by introducing postcursors. No significant difference between the standard condition and the off-frequency postcursor condition was found. This is indicated in the right panel which shows the release from BMLD due to the postcursors. However, there was a significant reduction of BMLD by the on-frequency postcursors (about 7 dB). Thus, the data demonstrates that high-frequency BMLD obtained with narrowband gaussian noise is affected by auditory grouping.

Fig. 4.5 shows the corresponding results obtained at low frequencies.

The amount of BMLD is larger than at high frequencies and the variability across subjects is much smaller. The effect of the on-frequency postcursors on the amount of BMLD is reduced (4 dB), indicating that BMLD is considerably less affected by

grouping at low frequencies than at high frequencies. This is consistent with a more "hard-wired" binaural processor based on temporal fine structure at low frequencies, while binaural envelope processing seems at least partly affected by the acoustical context.

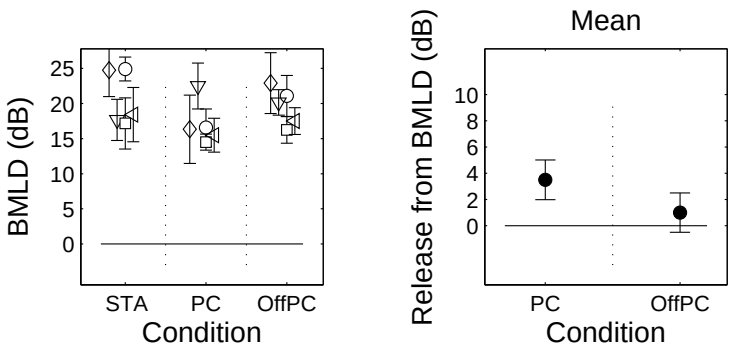

Figure 4.5: The left panel shows the low-frequency BMLD (signal frequency 500 Hz) for five subjects and three conditions ("STA", "PC", "OffPC") whereas in the right panel the release from unmaking for the five subjects by auditory stream segregation for two conditions ("PC", "OffPC") is shown.

In binaural envelope processing (BMLD) as well as in monaural across-frequency envelope processing (e.g., CMR), a processing circuit based on equalization and cancellation (EC) might be an effective strategy of "noise reduction" (e.g. see chapter 2 and Durlach, 1963). However, both circuits seem not to be completely "hard-wired" since they are influenced by grouping (7-8 dB) effect, CMR eliminated, BMLD reduced). The results when using transposed stimuli are shown in Fig. 4.6.

The amount of BMLD in the standard condition (STA) is slightly lower than when BMLD was measured with low-frequency masking bands. When postcursors are presented the BMLD decreases by around 4 db compared to the standard condition comparable to the situation for the low-frequency BMLD. The observation is consistent with the initial hypothesis that envelope processing is a more central process therefore sensitive to the auditory context (postcursors) while fine-structure of a stimuli is processed more peripherally. For the transposed stimuli the fine-structure

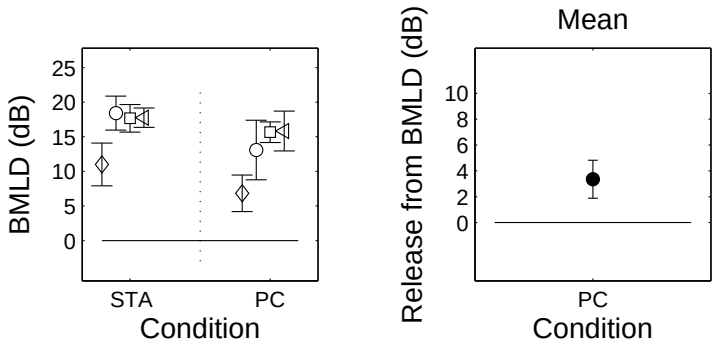

Figure 4.6: The left panel shows the amount of BMLD for the transposed stimuli (carrier frequency 5000 Hz) for four subjects anf three conditions ("STA", "PC") whereas in the right panel the release from unmaking for the five subjects by auditory stream segregation for one condition ("PC") is shown.

information is fully available although stimuli were presented at a frequency high enough to normally exclude all fine-structure processing.

4.5 Experiment 3: Modulation detection interference

4.5.1 Rationale

It was shown in the first two experiments that retroactive effects of perceptual segregation can lead to a deterioration in signal detection. To completely rule out any general interference effects that might lead to reduced detection performance, it would be most convincing to demonstrate that similar retroactive perceptual segregation can lead to enhanced detection in some situations. Yost et al. (e.g. 1989) observed that sinusoidal amplitude modulation (SAM) of a masker tone could increase modulation detection thresholds for a distant probe tone by as much as 15 dB if both tones were modulated at the same rate. They termed this increase of modulation thresholds in the presence of a modulated masker modulation detection interference (MDI). In contrast to CMR and BMLD, modulation detection interference (MDI) is an example of where across-frequency processing leads to a deterioration in performance. Thus, in this

case perceptual segregation of target and masker (interferer) might help in detecting the target modulation. In this part the impact of perceptual auditory grouping is investigated in a amplitude modulation paradigm.

4.5.2 Method

Apparatus and procedure

An adaptive, three-interval, three-alternative forced-choice procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point of the psychometric function. The signal modulation depth was set to -6 dB (in terms of 20 log m). The initial step size was 4 dB, which was reduced to 1 dB after four reversals. Threshold was defined as the mean of the levels at the last six reversals of a threshold run. Four threshold estimates were obtained and averaged from each listener in each condition.

Subjects

Six normal-hearing listeners participated in the experiments.

Stimuli

The stimuli were similar in configuration to those used in the CMR experiments, as shown in the upper row of Fig. 4.7. The signal was 16 Hz sinusoidal amplitude modulation imposed on a 1000-Hz pure-tone carrier, 187.5 ms in duration including 20-ms raised cosine ramps. The masker consisted of tonal carriers centered around frequencies of 250, 500, 1000, and 4000 Hz. The masker was either unmodulated (U) or amplitude modulated by a Gaussian noise with a center frequency of 16 Hz, cutoff frequencies of 8 and 24 Hz, and an rms level of -10 dB (re. unity). As in the CMR condition, the four postcursors (PM) at the masker frequencies were all the same duration as the target and the masker. They were separated by gaps of 62.5 ms, giving an overall repetition period of 250 ms. The condition PPM was introduced as a control experiment similar to those for the CMR experiment where the first of the four postcursors was omitted to destroy the "'auditory stream"' of flanking bands

and postcursors. The level of the target carrier was 60 dB SPL and the level of each flanking carrier was 54 dB SPL.

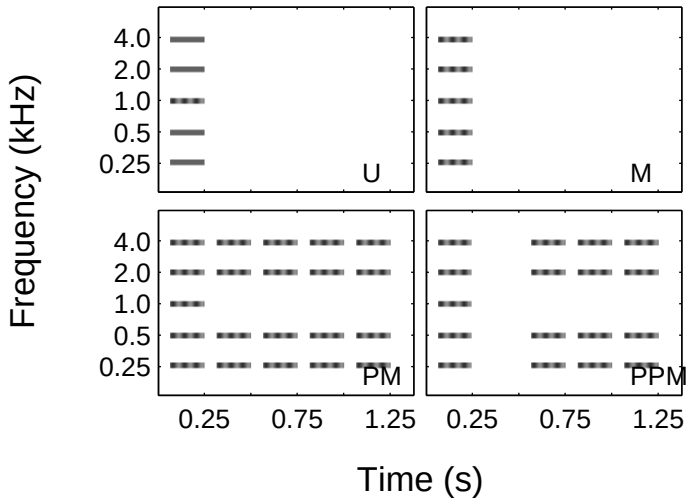

Figure 4.7: The panels show the four experimental conditions used in the MDI experiment: The unmodulated condition (U) where only the target band is modulated, the modulated condition (M) where all bands are comodulated, the modulated condition with postcursors (PM) and the modulated condition when the first postcursor was omitted.

4.5.3 Results

The results for the MDI experiment are shown in Fig. 4.8. In condition U when all the noise envelopes were uncorrelated the modulation detection threshold is about -14 dB relative to a fully amplitude modulated noise. In the classical MDI condition, the flanking bands were modulated (condition M), making detection of the target modulation more difficult with a modulation threshold rising to about -2 dB. The novel condition again involved postcursors, which were designed to capture the flanking bands and to perceptually segregate them from the target band (condition PM).

In this case, if across-frequency modulation processing is affected by perceptual

organisation, the segregation of the target from the flankers should increase the detectability of the target modulation leading to an improvement in performance. In fact modulation detection thresholds decrease again to approximately -12 dB. This is also the case for the condition (PPM) when the first postcursor was left out where modulation detection thresholds are found to be -11 dB thus similar to those found in condition U and PM.

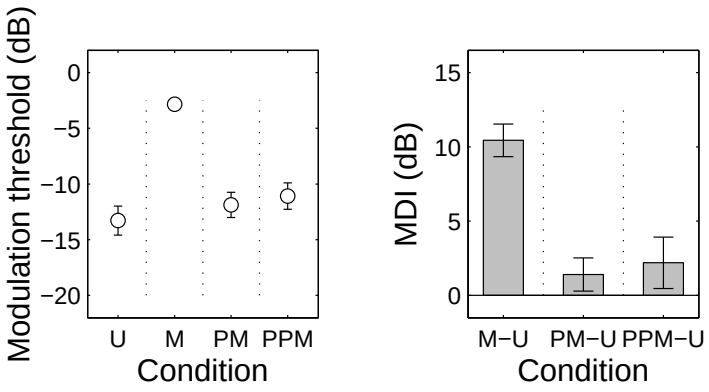

Figure 4.8: The left panel shows the modulation depth at threshold for the four conditions (U, M, PM, PPM) in Fig. 4.7 The right panel shows the amount of modulation detection interference (MDI), as the threshold difference M-U and PM-U, respectively.

The results of conditions U and M are similar to those of previous studies (Oxenham and Dau, 2001; Moore et al., 1995) of across-frequency modulation interference: modulation of the flanking bands resulted in a deterioration in the ability to detect the target modulation. The addition of the postcursors (condition PM) led to a complete elimination of the interference, so that thresholds were again at the same level as they were with the unmodulated flanking bands (no significant difference between conditions U and PM). Thus, so far the effects obtained with the comodulated flanking bands respectively postcursors have been diametrical for CMR and MDI, modulation detection itself is hindered by adding several flanker that share coherent envelopes with the target. In contrast when postcursors were added modulation detection became easier indicating that modulation can only be processed

efficiently across different frequency regions if these are grouped within the same auditory object or stream. However, other differences in the processing "strategies" in CMR versus MDI were found in additional experiments where the first of the four postcursors was left out in an otherwise unchanged experimental paradigm (condition PPM). While the data for CMR in experiment 1 (as well as for BMLD in experiment 2) showed the same results as for the corresponding condition without any postcursors (i.e. no perceptual segregation of signal and flanker bands), the amount of MDI was the same with and without the first postcursors. Thus, long-term adaptation due to the presentation of the modulated flankers from previous trials might have affected modulation detection (in MDI), while adaptation does not seem to play an important role in signal detection (CMR) in noise.

4.6 Discussion

Overall, the results show that across-frequency modulation processing does not act in isolation from the processes that give rise to auditory object and stream formation. Across-frequency modulation is necessary but not sufficient for the occurrence of both CMR and MDI. For CMR, our results clearly suggest high-level, rather than brainstem, interactions, whereby modulation can only be processed efficiently across-different frequency regions if these are grouped within the same auditory object or stream. Stated another way, the modulation analysis necessary for signal detection is performed on objects, rather than frequency channels. For BMLD, envelope-based across-ear processing is also (partially) affected by grouping, indicating that the binaural circuit is not completely hard-wired and that central processing is involved in the evaluation of interaural differences. In terms of MDI, the present study demonstrated that sound following the target completely eliminated the interference, suggesting similar conclusions about the effects of sequential streaming in this example of across-frequency modulation processing. Future studies need to show to what extent neural adaptation, which have been found to occur throughout the auditory pathway with multiple time constants (e.g. Ulanovsky et al., 2004), are involved in MDI.

The results of the present study might have important implications for speech

perception in noise. Amplitude modulations are known to be an important factor in speech perception (Shannon et al., 1995). The intelligibility of speech for normal-hearing listeners is better when speech is presented in a fluctuating rather than a continuous background noise when measured at the same signal to noise ratio. This decrease for in speech masking effectiveness due to the fluctuations of masking noise is known as masking release for speech. When compared to normal-hearing listeners, hearing-impaired listeners benefit less or not at all from the fluctuations in the masking noise, i.e. show reduced or no masking release for speech at all (Festen and Plomp, 1990). Masking release for speech is generally attributed to several factors, including the temporal resolution of the (normal) auditory system and CMR.

Interestingly, it has been shown that, after taking reduced audibility and nonlinear compression into account (Oxenham and Bacon, 2003), people with cochlear damage usually do not show impaired performance in certain temporal tasks like temporal integration function (Florentine et al., 1988), gap detection (Florentine and Buus, 1984; Nelson and Thomas, 1997), temporal modulation detection (Moore and Sek, 1992), and forward and backward masking (Nelson and Freyman, 1987). Still, many of these subjects seem to have problems in speech-in-noise understanding suggesting that other processing deficits might cause the difficulties. The current study demonstrated that across-channel modulation processing needs to be interpreted in terms of auditory object formation (and possibly in terms of long-term adaptation in the case of MDI). Thus, it is possible that many hearing-impaired listeners do have problems with speech perception in complex environments because their ability to perceptually organize sound is reduced and/or that their ability to detect sound (like speech) following previous stimulation (for example in an ongoing conversation) is impaired due to abnormal adaptation time constants in their processing. This needs to be investigated in future studies. The results of the present study place strong constraints on the search for neural correlates of across-frequency modulation processing. Interestingly, recent physiological studies demonstrated that neural activity in the primary auditory cortex (of the cat) represents sounds in terms of auditory objects rather than in terms of invariant acoustic features (Neuert et al., 2004; Las et al., 2005). Specifically, qualitative changes in the representation of tones in fluctuating noise were found along the auditory pathway, with a gradual segregation

of the tone from the noise towards higher centers in the auditory pathway. Based on their data, it was suggested that while most of the interesting auditory features and maps might already be extracted in the brainstem (e.g. the inferior colliculus), the organisation of these features into auditory objects takes place in the auditory cortex using temporal and spectral context at several time scales. The perceptual data from the present study seem to support this interpretation.

Finally, the results of this study also provide constraints on future models of modulation processing and perception. While the model from chapter 2, which reflects an across-channel extension of the original modulation filterbank model by Dau et al. (1997b), can account for a variety of detection and masking data, including across-channel CMR, it is not able to describe the nonlinear behavior observed in a varying acoustical context. Likewise, recent models of spectro-temporal processing in the auditory system by Shamma and colleagues (e.g. Chi et al., 2005), even though quite successful in the assessment of speech intelligibility, cannot account for the above findings. In both modeling approaches, the signal energy across time and frequency is essentially integrated linearly. This might be successful in relatively simple sound conditions but is limited in more complex sound situations. However, the output of the processing models, i.e., their internal representation, might provide some of the important auditory features as input to the "central processor". The models have been shown to be valuable as pre-processors in, for example, automatic speech recognition and objective assessment of speech quality (e.g. Hansen and Kollmeier, 1999; Tchorz and Kollmeier, 1999). Their value for the processing of complex stimuli and natural scenes is currently under investigation.

5

General discussion

This thesis investigated the phenomenon of comodulation masking release (CMR). CMR provides interesting insights into how the auditory system processes stimulus information that is distributed across frequency. Even though unnatural stimuli like tones and noises were used to study CMR, the results might be relevant for a better understanding of how the auditory system codes and analysis modulated stimuli such as speech and environmental sounds. The main focus of this work has been the development of a computational model of auditory signal processing that can account for many key observation in the experimental data from a broad range of experimental paradigms. It was attempted to provide a framework that allows to discriminate between the contributions to CMR from the different stages of processing along the auditory pathway. Specifically, the goal was to provide a model that covers "automatically" and simultaneously both across-frequency processes as well as within-channel processes in the same framework, whereby the relative contributions of within versus across-channel processing depend on the specific stimulus characteristics and the level and frequency dependent properties of the auditory system. In previous investigations on CMR, the modeling has typically been at a rather qualitative level or focused on only one particular aspect of CMR, such as the characteristics of either within-channel or across-channel processing, or on the effect of nonlinear cochlear processing on CMR. It remained unclear, which aspects in the data could be attributed to which underlying mechanism, particularly in experimental where most likely more than one process was involved.

In Chapter 2, an explicit across-channel processing was suggested and implemented into an existing monaural model of auditory signal processing (Dau et al., 1996, 1997a). The across-channel process, that accounted for "true" CMR, was motivated and based on the equalization-cancellation (EC) mechanism which is well-

known from models of across-ear, i.e. binaural, processing (e.g. Durlach, 1963). The EC process is closely related to a correlation analysis across channels. In the CMR model presented here, the internal activity of the stimuli at the output of individual modulation filters is integrated across all peripheral filters, or frequency channels, except for the filter at the signal frequency. This summed activity was averaged (representing the E-process) and subtracted from the outputs of the modulation filters belonging to the peripheral filter centered at the signal frequency (representing the C process). The model was able to account for the data of several across- and within-channel CMR paradigms. Specifically, the model predicted an across-channel CMR effect that was maximally 3-5 dB, even in cases when several remote flanking bands were considered. This suggested that in all conditions where a large CMR was found, most of this should be attributed to within-channel processing. Furthermore, conceptually, the results suggested that a modeling approach based on cross-correlation appears to be successful not only for the processing of sound across ears but also for the processing of modulations across frequency channels within the same ear.

In Chapter 3, the processing model was further generalized by introducing a functional nonlinear peripheral filtering stage in form of the dual resonance non-linear filterbank (DRNL Meddis et al., 2001). This stage introduced peripheral compression which is crucial for the correct prediction of various masking data (e.g. Oxenham and Plack, 1997; Ernst and Verhey, 2006; Jepsen et al., 2008) as well as aspects associated with suppression (Meddis et al., 2001). The inclusion of the level dependent peripheral filtering stage allowed a detailed analysis of level effects in CMR, specifically the dependency of CMR on the relative level of the masker components, the overall masker level and the number of masking components. The model was evaluated by comparing the predictions with results either from newly designed experiments or from the literature. In addition to the modifications in the peripheral stage, the model was also generalized in terms of an automatic process that determines the channels involved in the EC process, in contrast to the earlier simulations where the signal frequency was one of the parameters that had to be provided as input information. The results from the simulations in this chapter suggested that there are (at least) three major process/components that contribute to the overall amount of CMR: (i) within-

channel beating, corresponding to the results from Verhey et al. (1999), (ii) within-channel suppression, accounted for by the extended model but not the model from chapter 2, and (iii) explicit across-channel processing as proposed in chapter 2. Some of the predictions were quite sensitive to slight parameter changes particularly in the cochlear processing stage, such as the exact amount of assumed compression and the kneepoint between linear and compressive transformation in the cochlear input-output function. Some of these dependencies need to be investigated further in order to better understand the exact relations. Overall, however, the proposed model was able to account for a very large number of conditions using the same framework.

Chapter 4 investigated the effects of auditory grouping on CMR. It was found that perceptual grouping largely affected results in conditions of across-channel processing. CMR was fully eliminated by presenting a group of "postcursors" after the flanking bands (but not the signal band). In such a condition, the flanking bands and the postcursers were perceptually grouped together and thereby segregated from the signal band, i.e. the comparison across frequency which normally produces CMR did not take place. In contrast, when the flanking bands and postcursors were placed close to each other in frequency, no grouping or context effects were found. These results supported the hypothesis that there is an across-channel component of CMR, susceptible to grouping manipulations and probably taking place at a central stage of processing, and a within-channel component of CMR, which does not depend on variations of the acoustical context and it probably coded more peripherally in the processing. In order to further investigate the role of grouping on across-channel modulation processing, another across-channel effect, modulation detection interference (MDI) was considered and also found to be highly dependent on manipulations of auditory grouping, similarly as CMR. It may be a general auditory organizational principle that access to higher-order percepts, such as modulation strength, is generally afforded only after the outputs of peripheral frequency channels are combined to form auditory objects or streams. In other words, access to sensations is via objects rather than outputs of peripheral auditory channels. Finally, also binaural level differences were considered in the context of auditory grouping. It was found that envelope-based BMLD at high (audio) frequencies was susceptible to perceptual auditory grouping. BMLD was reduced by about 6 dB, but not eliminated, when

postcursors suceeded the on-frequency masker at one ear, compared to the "standard" BMLD condition without any context information. The size of this reduction was similar to that found in across-channel CMR. It is possible that the mechanisms underlying envelope-based BMLD and across-channel CMR are similar, consistent with results from van de Par (1998), and that both processes take place at a more central stage of processing. In contrast, fine-structure based BMLD at low (audio) frequencies was not affected by the postcursors which suggests that this processing is coded at a more peripheral stage of auditory processing.

In terms of modeling, the grouping results on CMR cannot be accounted for by the model developed in this thesis. Specifically, it is not able to simulate the elimination of CMR as a result of perceptual segregation of the masker band from the flanker bands. Likewise, other recent models of spectro-temporal processing in the auditory system, proposed by Shamma and colleagues (e.g. Chi et al., 2005), cannot account for this finding. In both modeling approaches, the signal energy across time and frequency is essentially integrated linearly. This might be successful in relatively "simple" sound conditions but fails in more complex sound situations where perception depends on the acoustical context. However, the output of the processing models might provide some of the important auditory features as input to the "central processor". In a very recent study, Elhilali et al. (2009) suggested a model of auditory streaming based on temporal coherence between different sounds features (such as different frequency channels). They suggested the analysis of a dynamic coherence matrix between each pair of channels and proposed an Eigenvalue analysis to decompose each matrix into coherent components and determine the number of independent streams. Interestingly, their analysis was also performed at the output of a temporal modulation filterbank (after cochlear preprocessing), with characteristics similar to that used in the present study. Also the correlation network proposed in Elhilali et al. (2009) is conceptually similar to the correlation analysis carried out through the EC network of the model from the present study. Thus, it would be interesting to combine the concepts of the two studies to test to what extent grouping effects in CMR can be accounted for quantitatively. Furthermore, such a combined analysis could be interesting for investigating the auditory processing and perception of more general stimuli including speech and environmental sounds.

Bibliography

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (1971). Properties of 'two-tone inhibition' in primary auditory neurones. *J Physiol*, **212**(3), 593–609.

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (1997). Masking by modulated and unmodulated noise: effects of bandwidth, modulation rate, signal frequency, and masker level. *J Acoust Soc Am*, **101**(3), 1600–1610.

Berg, B. G. (1996). On the relation between comodulation masking release and temporal modulation transfer functions. *J Acoust Soc Am*, **100**(2 Pt 1), 1013–1023.

Bernstein, L. R. and Trahiotis, C. (1996). The normalized correlation: accounting for binaural detection across center frequency. *J Acoust Soc Am*, **100**(6), 3774–3784.

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**(2), 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001b). Binaural processing model based on contralateral inhibition. ii. dependence on spectral parameters. *J Acoust Soc Am*, **110**(2), 1089–1104.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001c). Binaural processing model based on contralateral inhibition. iii. dependence on temporal parameters. *J Acoust Soc Am*, **110**(2), 1105–1117.

Bregman, A. S., Liao, C., and Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Can J Psychol*, **44**(3), 400–413.

Buschermöhle, M., Verhey, J. L., Feudel, U., and Freund, J. A. (2007). The role of the auditory periphery in comodulation detection difference and comodulation masking release. *Biol Cybern*, **97**(5-6), 397–411.

Buss, E. and Hall, J. W. (1998). The role of auditory filters in comodulation masking release (cmr). *J Acoust Soc Am*, **103**(6), 3561–3566.

Buss, E., Hall, J. W., and Grose, J. H. (1998). Change in envelope beats as a possible cue in comodulation masking release (cmr). *J Acoust Soc Am*, **103**(3), 1592–1597.

Buus, S. (1985). Release from masking caused by envelope fluctuations. *J Acoust Soc Am*, **78**(6), 1958–1965.

Carlyon, R. and Stubbs, R. (1989). Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers. *Journal of the Acoustical Society of America*, **85**(6), 2563–2574.

Chi, T., Ru, P., and Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am*, **118**(2), 887–906.

Cohen, M. (1991). Comodulation masking release over a 3 octave range. *J Acoust Soc Am*, **90**(4 Pt 1), 1381–1384.

Cohen, M. F. and Schubert, E. D. (1987). The effect of cross-spectrum correlation on the detectability of a noise band. *J Acoust Soc Am*, **81**(3), 721–723.

Dau, T. (1996). *Modeling auditory processing of amplitude modulation*. PhD thesis, Universität Oldenburg.

Dau, T., Ewert, S., and Oxenham, A. J. (2009). Auditory stream formation affects comodulation masking release retroactively. *J Acoust Soc Am*, **125**(4), 2182–2188.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *J Acoust Soc Am*, **102**(5 Pt 1), 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration. *J Acoust Soc Am*, **102**(5 Pt 1), 2906–2919.

Dau, T., Püschel, D., and Kohlrausch, A. (1996). A quantitative model of the "effective" signal processing in the auditory system. i. model structure. *J Acoust Soc Am*, **99**(6), 3615–3622.

Dau, T., Verhey, J., and Kohlrausch, A. (1999). Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. *J Acoust Soc Am*, **106**(5), 2752–2760.

Derleth, R. P. and Dau, T. (2000). On the role of envelope fluctuation processing in spectral masking. *J Acoust Soc Am*, **108**(1), 285–296.

Derleth, R. P., Dau, T., and Kollmeier, B. (2001). Modeling temporal and compressive properties of the normal and impaired auditory system. *Hear Res*, **159**(1-2), 132–149.

Domnitz, R. H. and Colburn, H. S. (1977). Lateral position and interaural discrimination. *J Acoust Soc Am*, **61**(6), 1586–1598.

Duifhuis, H. (1980). Level effects in psychophysical two-tone suppression. *J Acoust Soc Am*, **67**(3), 914–927.

Durlach, N. (1960). Note on the equalization and cancellation theory of binaural masking level differences. *J Acoust Soc Am*, **32**, 1075–1076.

Durlach, N. (1963). Equalization and cancellation theory of binaural masking-level differences. *J Acoust Soc Am*, **35**(8), 1206–1218.

Eddins, D. A. (2001). Measurement of auditory temporal processing using modified masking period patterns. *J Acoust Soc Am*, **109**(4), 1550–1558.

Eddins, D. A. and Wright, B. A. (1994). Comodulation masking release for single and multiple rates of envelope fluctuation. *J Acoust Soc Am*, **96**(6), 3432–3442.

Elhilali, M., Xiang, J., Shamma, S. A., and Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol*, **7**(6), e1000129.

Ernst, S. M. A. and Verhey, J. L. (2005). Comodulation masking release over a three octave range. *J Acoust Soc Am*, **91**(6), 998–1006.

Ernst, S. M. A. and Verhey, J. L. (2006). Role of suppression and retro-cochlear processes in comodulation masking release. *J Acoust Soc Am*, **120**(6), 3843–3852.

Ewert, S. D. and Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. *J Acoust Soc Am*, **108**(3 Pt 1), 1181–1196.

Ewert, S. D. and Dau, T. (2004). External and internal limitations in amplitude-modulation processing. *J Acoust Soc Am*, **116**(1), 478–490.

Ewert, S. D., Verhey, J. L., and Dau, T. (2002). Spectro-temporal processing in the envelope-frequency domain. *J Acoust Soc Am*, **112**(6), 2921–2931.

Festen, J. M. and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am*, **88**(4), 1725–1736.

Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.*, **12**(1), 47–65.

Florentine, M. and Buus, S. (1984). Temporal gap detection in sensorineural and simulated hearing impairments. *J Speech Hear Res*, **27**(3), 449–455.

Florentine, M., Fastl, H., and Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *J Acoust Soc Am*, **84**(1), 195–203.

Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. Wiley, New York.

Green, D. M. (1992). On the similarity of two theories of comodulation masking release. *J Acoust Soc Am*, **91**(3), 1769.

Grose, J. H. and Hall, J. W. (1989). Comodulation masking release using sam tonal complex maskers: effects of modulation depth and signal position. *J Acoust Soc Am*, **85**(3), 1276–1284.

Grose, J. H. and Hall, J. W. (1993). Comodulation masking release: is comodulation sufficient? *J Acoust Soc Am*, **93**(5), 2896–2902.

Haggard, M., Harvey, A., and Carlyon, R. (1985). Peripheral and central components of comodulation masking release. *J Acoust Soc Am*, **78**, 63–63.

Haggard, M. P., Hall, J. W., and Grose, J. H. (1990). Comodulation masking release as a function of bandwidth and test frequency. *J Acoust Soc Am*, **88**(1), 113–118.

Hall, J. W. (1986). The effect of across-frequency differences in masking level on spectro-temporal pattern analysis. *J Acoust Soc Am*, **79**(3), 781–787.

Hall, J. W., Grose, J. H., and Haggard, M. P. (1990). Effects of flanking band proximity, number, and modulation pattern on comodulation masking release. *J Acoust Soc Am*, **87**(1), 269–283.

Hall, J. W., Grose, J. H., and Hatch, D. R. (1996). Effects of masker gating for signal detection in unmodulated and modulated bandlimited noise. *J Acoust Soc Am*, **100**(4 Pt 1), 2365–2372.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). Detection in noise by spectro-temporal pattern analysis. *J Acoust Soc Am*, **76**(1), 50–56.

Hansen, M. and Kollmeier, B. (1999). Continuous assessment of time-varying speech quality. *J Acoust Soc Am*, **106**(5), 2888–2899.

Hatch, D., Arne, B., and Hall, J. (1995). Comodulation masking release (cmr): effects of gating as a function of number of flanking bands and masker bandwidth. *J Acoust Soc Am*, **97**(6), 3768–3774.

Hicks, M. L. and Bacon, S. P. (1995). Some factors influencing comodulation masking release and across-channel masking. *J Acoust Soc Am*, **98**(5 Pt 1), 2504–2514.

Hohmann, V. and Kollmeier, B. (2007). *A Nonlinear Auditory Filterbank Controlled by Sub-band Instantaneous Frequency Estimates*, chapter 3.

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am*, **51**(6), 1885–1894.

Houtgast, T. (1974). *Lateral suppression in hearing*. PhD thesis, Free University of Amsterdam.

Irino, T. and Patterson, R. D. (1997). A time-domain, level-dependent auditory filter: The gammachirp. *J Acoust Soc Am*, **101**(1), 412–419.

Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). A computational model of human auditory signal processing and perception. *J Acoust Soc Am*, **124**(1), 422–438.

Klump, G. M. and Langemann, U. (1995). Comodulation masking release in a songbird. *Hear Res*, **87**(1-2), 157–164.

Langemann, U. and Klump, G. M. (2001). Signal detection in amplitude-modulated maskers. i. behavioural auditory thresholds in a songbird. *Eur J Neurosci*, **13**(5), 1025–1032.

Langner, G. (1992). Periodicity coding in the auditory system. *Hear Res*, **60**(2), 115–142.

Las, L., Stern, E. A., and Nelken, I. (2005). Representation of tone in fluctuating maskers in the ascending auditory system. *J Neurosci*, **25**(6), 1503–1513.

Lawson, J. and Uhlenbeck, G. (1950). *Threshold signals*, volume 24 of *Radiation Laboratory*. McGraw Hill Book Co.

Lee, J. and Bacon, S. P. (1997). Amplitude modulation depth discrimination of a sinusoidal carrier: effect of stimulus duration. *J Acoust Soc Am*, **101**(6), 3688–3693.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, **49**(2), Suppl 2:467+.

Lutfi, R. A. and Patterson, R. D. (1984). On the growth of masking asymmetry with stimulus intensity. *J Acoust Soc Am*, **76**(3), 739–745.

McFadden, D. (1986). Comodulation masking release: effects of varying the level, duration, and time delay of the cue band. *J Acoust Soc Am*, **80**(6), 1658–1667.

McFadden, D. (1987). Comodulation detection differences using noise-band signals. *J Acoust Soc Am*, **81**(5), 1519–1527.

Meddis, R., Delahaye, R., O'Mard, L., Summer, C., Fantini, D. A., Winter, I., and Pressnitzer, D. (2002). A model of signal processing in the cochlear nucleus: Comodulation masking release. *Acta Acustica (Stuttgart)*, **88**(3), 387–398.

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**(6), 2852–2861.

Moore, B. C. and Emmerich, D. S. (1990). Monaural envelope correlation perception, revisited: effects of bandwidth, frequency separation, duration, and relative level of the noise bands. *J Acoust Soc Am*, **87**(6), 2628–2633.

Moore, B. C. and Glasberg, B. R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear Res*, **28**(2-3), 209–225.

Moore, B. C., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). The shape of the ear's temporal window. *J Acoust Soc Am*, **83**(3), 1102–1116.

Moore, B. C., Hall, J. W., Grose, J. H., and Schooneveldt, G. P. (1990). Some factors affecting the magnitude of comodulation masking release. *J Acoust Soc Am*, **88**(4), 1694–1702.

Moore, B. C. and Sek, A. (1992). Detection of combined frequency and amplitude modulation. *J Acoust Soc Am*, **92**(6), 3119–3131.

Moore, B. C., Sek, A., and Shailer, M. J. (1995). Modulation discrimination interference for narrow-band noise modulators. *J Acoust Soc Am*, **97**(4), 2493–2497.

Moore, B. C. and Shailer, M. J. (1991). Comodulation masking release as a function of level. *J Acoust Soc Am*, **90**(2 Pt 1), 829–835.

Mott, J. B., McDonald, L. P., and Sinex, D. G. (1990). Neural correlates of psychophysical release from masking. *J Acoust Soc Am*, **88**(6), 2682–2691.

Nelken, I., Rotman, Y., and Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*, **397**(6715), 154–157.

Nelson, D. A. and Freyman, R. L. (1987). Temporal resolution in sensorineural hearing-impaired listeners. *J Acoust Soc Am*, **81**(3), 709–720.

Nelson, P. B. and Thomas, S. D. (1997). Gap detection as a function of stimulus loudness for listeners with and without hearing loss. *J Speech Lang Hear Res*, **40**(6), 1387–1394.

Neuert, V., Verhey, J. L., and Winter, I. M. (2004). Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers. *J Neurosci*, **24**(25), 5789–5797.

Nieder, A. and Klump, G. M. (2001). Signal detection in amplitude-modulated maskers. ii. processing in the songbird's auditory forebrain. *Eur J Neurosci*, **13**(5), 1033–1044.

Oxenham, A. J. and Bacon, S. P. (2003). Cochlear compression: perceptual measures and implications for normal and impaired hearing. *Ear Hear*, **24**(5), 352–366.

Oxenham, A. J. and Dau, T. (2001). Modulation detection interference: effects of concurrent and sequential streaming. *J Acoust Soc Am*, **110**(1), 402–408.

Oxenham, A. J. and Moore, B. C. (1994). Modeling the additivity of nonsimultaneous masking. *Hearing Research*, **80**(1), 105–118.

Oxenham, A. J. and Plack, C. J. (1997). A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *J Acoust Soc Am*, **101**(6), 3666–3675.

Palmer, A. and Russell, I. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, **24**, 1–15.

Patterson, R. and Moore, B. (1986). Auditory filters and excitation patterns as representations of frequency resolution. *Frequency Selectivity in Hearing*, 123–178.

Patterson, R. D., Nimmo-Smith, J., Holdsworth, J., and Rice, P. (1987). An efficient auditory filterbank based on the gammatone function. *Paper presented at a meeting of the IOC Speech Group on Auditory Modeling at RSRE, December 14-15*.

Piechowiak, T., Ewert, S. D., and Dau, T. (2007). Modeling comodulation masking release using an equalization-cancellation mechanism. *J Acoust Soc Am*, **121**(4), 2111–2126.

Piechowiak, T., Ewert, S. D., and Dau, T. (2009). Modeling cmr: A unifying approach. *J Acoust Soc Am*. submitted.

Plack, C. J., Oxenham, A. J., and V.Dgra (2002). Linear and nonlinear processes in temporal masking. *Acta Acoustica united with Acoustica*, **88**, 348 – 358.

Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (2001). Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. *J Neurosci*, **21**(16), 6377–6386.

Püschel, D. (1988). *Prinzipien der zeitlichen Analyse beim Hören*. PhD thesis, Universität Göttingen.

Rhode, W. S. and Robles, L. (1974). Evidence from mössbauer experiments for nonlinear vibration in the cochlea. *J Acoust Soc Am*, **55**(3), 588–596.

Richards, V. M. (1987). Monaural envelope correlation perception. *J Acoust Soc Am*, **82**(5), 1621–1630.

Robles, L. and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiol Rev*, **81**(3), 1305–1352.

Rotman, Y., Bar-Yosef, O., and Nelken, I. (2001). Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. *Hear Res*, **152**(1-2), 110–127.

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**(4), 2151–2163.

Ruggero, M. A., Robles, L., and Rich, N. C. (1992). Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. *J Neurophysiol*, **68**(4), 1087–1099.

Sachs, M. B. and Kiang, N. Y. (1968). Two-tone inhibition in auditory-nerve fibers. *J Acoust Soc Am*, **43**(5), 1120–1128.

Schooneveldt, G. P. and Moore, B. C. (1987). Comodulation masking release (cmr): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band. *J Acoust Soc Am*, **82**(6), 1944–1956.

Schooneveldt, G. P. and Moore, B. C. (1989). Comodulation masking release (cmr) as a function of masker bandwidth, modulator bandwidth, and signal duration. *J Acoust Soc Am*, **85**(1), 273–281.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, **270**(5234), 303–304.

Snedecor, G. and Cochran, W. (1989). *Statistical Methods*. Iowa State University Press; 8 edition.

Tchorz, J. and Kollmeier, B. (1999). A model of auditory perception as front end for automatic speech recognition. *J Acoust Soc Am*, **106**(4 Pt 1), 2040–2050.

Thompson, E. R. and Dau, T. (2008). Binaural processing of modulated interaural level differences. *J Acoust Soc Am*, **123**(2), 1017–1029.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*, **24**(46), 10440–10453.

van de Par, S. (1998). *A comparison of binaural detection at low and high frequencies*. PhD thesis, Technische Universiteit Eindhoven.

van de Par, S. and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. *J Acoust Soc Am*, **101**(3), 1671–1680.

van de Par, S. and Kohlrausch, A. (1998). Diotic and dichotic detection using multiplied-noise maskers. *J Acoust Soc Am*, **103**(4), 2100–2110.

Verhey, J. L. (2002). Modeling the influence of inherent envelope fluctuations in simultaneous masking experiments. *J Acoust Soc Am*, **111**(2), 1018–1025.

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**(5), 2733–2745.

Verhey, J. L., Pressnitzer, D., and Winter, I. M. (2003). The psychophysics and physiology of comodulation masking release. *Exp Brain Res*, **153**(4), 405–417.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am*, **66**(5), 1364–1380.

Viemeister, N. F. and Wakefield, G. H. (1991). Temporal integration and multiple looks. *J Acoust Soc Am*, **90**(2 Pt 1), 858–865.

Yates, G. K., Winter, I. M., and Robertson, D. (1990). Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hear Res*, **45**(3), 203–219.

Yost, W. A., Sheft, S., and Opie, J. (1989). Modulation interference in detection and discrimination of amplitude modulation. *J Acoust Soc Am*, **86**(6), 2138–2147.