

Neurosensory Modeling of Signal Detection in Comodulated Noise

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Dipl.-Phys. Michael Buschermöhle
geboren am 09. Juni 1977
in Wildeshausen

Gutachterin: Prof. Dr. Ulrike Feudel
Zweitgutachter: Juniorprof. Dr. Jesko L. Verhey
Tag der Disputation: 16.07.2007

ABSTRACT

Many sounds in our everyday experience exhibit a property called comodulation: they have common amplitude modulations in different frequency regions. In various situations, these comodulated sounds hamper the understanding or even detection of relevant sounds, thus acting as noise maskers for the signals of interest. In the scientific literature, signal detection in comodulated noise has been examined with two experimental paradigms, comodulation masking release (CMR) and comodulation detection difference (CDD). The present dissertation is concerned with modeling the processes underlying these experiments on a neurosensory basis. This means that models are derived which are based on neurophysiological knowledge of the auditory system. The aim of the present work is to focus on parsimonious models which preferably cover a wide range of experiments and which can be examined mostly with analytical methods. By thoroughly examining these models the contributions of different peripheral sensory mechanisms to the abovementioned comodulation effects can be estimated.

It is known from investigations of CMR and CDD experiments that humans and other animals can have lower signal detection thresholds if auditory signals are embedded in comodulated background noise than if incoherently modulated or unmodulated background noise is present. During the course of this thesis, this effect is shown to have two important peripheral contributions. The first one is the extraction of the envelope of the appropriately filtered stimulus, which happens in the inner ear when basilar membrane vibrations are encoded in auditory nerve action potentials. The second contribution is a compressive nonlinearity which is applied to the basilar membrane vibrations by active mechanisms in the inner ear.

Due to the fact that comodulation does not only occur in the realm of air pressure vibrations but also in any other medium for which vibrations are possible, the results of the present thesis may be extended to other areas like e.g. electromagnetic waves or ultrasound.

ZUSAMMENFASSUNG

Viele Geräusche aus unserer alltäglichen Umwelt haben eine Eigenschaft, die als Komodulation bezeichnet wird: Sie haben gleichzeitige Amplitudenmodulationen in verschiedenen Frequenzbereichen. In vielen Situationen erschweren diese komodulierten Geräusche die Verständlichkeit oder sogar die Detektierbarkeit wichtiger Signale und wirken daher als Maskierer. In der wissenschaftlichen Literatur ist Signaldetektion in komoduliertem Rauschen mit zwei experimentellen Paradigmen untersucht worden, *comodulation masking release* (CMR) und *comodulation detection difference* (CDD). Die vorliegende Dissertation beschäftigt sich auf Basis der Neurosensorik mit der Modellierung der Prozesse, die diesen Experimenten zugrunde liegen. Das bedeutet, dass Modelle hergeleitet werden, die auf neurophysiologischem Wissen über das auditorische System beruhen. Das Ziel der vorliegenden Arbeit ist es, sich auf Modelle zu konzentrieren, die vorzugsweise eine breite Spanne von Experimenten abdecken und die größtenteils mit analytischen Methoden untersucht werden können. Durch eingehende Untersuchung dieser Modelle können die Beiträge verschiedener peripherer sensorischer Mechanismen auf die oben genannten Komodulationseffekte abgeschätzt werden.

Aus Untersuchungen von CMR- und CDD-Experimenten ist bekannt, dass Menschen und andere Tiere niedrigere Signaldetektionsschwellen haben können, falls die auditorischen Signale in komoduliertes Hintergrundrauschen eingebettet sind als in Gegenwart von inkohärent moduliertem oder unmoduliertem Hintergrundrauschen. Im Verlauf der Dissertation wird gezeigt, dass dieser Effekt zwei wichtige periphere Beiträge hat. Der erste ist die Extraktion der Einhüllenden des geeignet gefilterten Stimulus, die im Innenohr geschieht, wenn Basilarmembranschwingungen in Aktionspotentiale des Auditorischen Nervs übersetzt werden. Der zweite Beitrag ist eine kompressive Nichtlinearität, der die Basilarmembranschwingungen aufgrund von aktiven Mechanismen im Innenohr ausgesetzt sind.

Da Komodulation nicht nur im Bereich der Luftdruckschwingungen vorkommen kann, sondern auch in jedem anderen vibrationsfähigen Medium, können die Ergebnisse dieser Dissertation auf andere Bereiche wie z.B. elektromagnetische Wellen oder Ultraschall ausgedehnt werden.

CONTENTS

1	Introduction	1
1.1	Overview	10
2	Signal Detection in Comodulated Noise	13
2.1	Abstract	13
2.2	Introduction	14
2.3	Experimental background	16
2.4	Analytically tractable model	17
2.4.1	Model description	17
2.4.2	Model results	22
2.5	Enhanced model with numerical simulations	24
2.5.1	Model processing steps	24
2.5.2	Results of model simulations	26
2.5.3	Discussion	29
2.6	Conclusions	31
2.7	Appendix: Firing rate standard deviations	32
2.8	Appendix: Influence of filter bandwidth and compression	34
3	A Unifying Model Approach to Comodulation Detection Difference (CDD) and Comodulation Masking Release (CMR)	37
3.1	Abstract	37
3.2	Introduction	37
3.3	Model structure	42
3.3.1	Mechanism of signal detection	47
3.3.2	Model parameters	47
3.4	Results	48
3.4.1	Frequency spacing between signal and flanking bands for CDD	50
3.4.2	Frequency spacing between the signal and one flanking band for CMR	51
3.4.3	Number of flanking bands for CDD and CMR	52
3.4.4	Influence of noise band level	56
3.5	Discussion	58
3.5.1	Filter bandwidth γ	58

3.5.2	Compression α	59
3.5.3	Signal detection criterion D and internal error σ	59
3.5.4	Parameter dependence of the model results	60
3.5.5	Influence of temporal processing	60
3.6	Conclusions	62
3.7	Appendix: Envelope statistics with levels	62
4	Enhanced Signal Detectability in Comodulated Noise Introduced by Compression	65
4.1	Abstract	65
4.2	Introduction	65
4.3	Properties of comodulated noise	67
4.4	Envelope compression	69
4.4.1	Influence of compression on the standard deviation of unmodulated noise	71
4.4.2	Influence of compression on the standard deviation of comodulated noise	72
4.5	Signal detection in (un)compressed noise	73
4.5.1	SNR for narrowband noise signals (CDD)	75
4.5.2	SNR for pure tone signals (CMR)	76
4.5.3	Improved signal detectability in comodulated noise after compression	76
4.5.4	Causes of signal detection differences	79
4.5.5	Example: instantaneous compression	81
4.6	Discussion and conclusions	82
4.7	Appendix: Envelope distribution of comodulated noise	84
4.8	Appendix: Calculations of the SNR	84
4.8.1	Approximations for small narrowband signals (CDD)	84
4.8.2	Approximations for large pure tone signals (CMR)	86
4.8.3	Approximations for large narrowband signals and small pure tone signals	87
5	Summary and concluding remarks	89
5.1	Summary of Chapters 2 to 4	89
5.2	General conclusions	92
5.3	Possibilities for future research	93
	Bibliography	97
	Acknowledgments	105

Curriculum Vitae

107

CHAPTER 1

INTRODUCTION

The human auditory system is an extraordinarily powerful organ: it can adapt to extreme situations and detect sonic signals under various adverse circumstances. To be more precise, signals are not only detected but also classified, separated, and occasionally attended to. Usually, a stimulus reaching the ear may be said to consist of several so-called auditory objects (Griffiths and Warren, 2004), which refers to the fact that sounds originating from different sources or having certain distinguishing properties can be perceptually separated while other sounds may be perceptually combined. The process of disentangling the mixture of sounds arriving at the ear has been termed auditory scene analysis (e.g. Bregman, 1999). One example where auditory scene analysis proves necessary is a conversation in a crowded social environment which can be followed despite strong background noise. This has been described as the cocktail party effect; see Arons (1992) and Bronkhorst (2000) for reviews.

The act of detecting signals in background noise may be viewed as a very basic form of auditory scene analysis. After millions of years of evolution (e.g. Manley et al., 2004), the auditory systems of humans and other animals have developed amazing capabilities of sensing the presence of possibly meaningful new sounds in ongoing background clutter. For example, it is vitally important for prey like antelopes to detect the rustling of shrubbery caused by predators like lions sneaking up to them through the brushwood in the African savanna. Different cues may guide the auditory system when performing such signal detection tasks. For example, the spectral content of a new sound may be different from that of the ongoing sound mixture. Or the auditory system may take advantage of the general spectro-temporal structure of everyday sounds and signals when performing signal detection tasks (e.g. Rieke et al., 1995; Smith and Lewicki, 2006).

It has been pointed out by Nelken et al. (1999) that many natural sounds have a property that has been termed comodulation: they exhibit common amplitude fluctuations at many frequency regions. This is in contrast to sounds with essentially random amplitude fluctuations across the spectrum (e.g. if many speakers with different voices in one room talk at the same time)

or no noteworthy amplitude fluctuations at all (e.g. pure tones or white noise). Common amplitude modulation may arise, for example, when sounds travel towards the detector through turbulent air (Richards and Wiley, 1980; Wiley and Richards, 1982; Embleton, 1996). Even more prevalent may be the origin of comodulation from the process of sound generation by the vocal tracts of animals (Singh and Theunissen, 2003). Due to the ubiquitous presence of comodulated sounds in natural habitats, it is likely that animals have adapted in order to cope with the specific features of comodulated sounds (cf. Klump, 1996; Brumm and Slabbekoorn, 2005). Indeed, it has been shown that humans and other animals can be better at detecting signals in comodulated background noise than in unmodulated noise. Two prominent experimental paradigms address the detection of signals in comodulated noise: *comodulation masking release* (CMR), which was introduced by Hall et al. (1984), and *comodulation detection difference* (CDD), which was first described by McFadden (1987) and Cohen and Schubert (1987). The fundamental ideas of these two experiments will be illustrated in the following.

Hall et al. (1984) demonstrated that human detection thresholds for sine tones masked by comodulated noise are lower than in unmodulated noise. These thresholds are called masked thresholds because the background noise prevents the tone from being detected at absolute threshold and therefore masks the presence of the tone signal. Hence, the background noise may be called a masking noise or masker. In the abovementioned study, the difference between detection thresholds of a sine tone in comodulated noise and in unmodulated noise has been termed the *comodulation masking release* (CMR), because the amount of masking was reduced when the comodulated masker was present. Exemplary CMR stimuli are sketched in the top row of Fig. 1.1. Subsequent studies have investigated this effect in depth in human psychophysics. Examples are Schooneveldt and Moore (1987, 1989); Verhey et al. (2003), and Ernst and Verhey (2006). Animal studies on CMR have been performed in terms of behavior (Langemann and Klump, 2001) and neurophysiology (Nelken et al., 1999; Hofer and Klump, 2003), where an effect of similar size to that in humans was found.

The second experimental paradigm, CDD, investigates the detectability of narrow band noise signals instead of pure tones. In the corresponding experiments the masking noise is usually constructed from several narrow noise bands that can share the same amplitude fluctuations (comodulated maskers) or have independent amplitude fluctuations (uncorrelated maskers). Exemplary stimuli are sketched in the bottom row of Fig. 1.1. One finds reduced signal detection thresholds if the signal band has to be detected in comodulated masker bands (the co-uncorrelated condition in the bottom row of Fig. 1.1) compared to the situation of uncorrelated maskers (the all uncorrelated

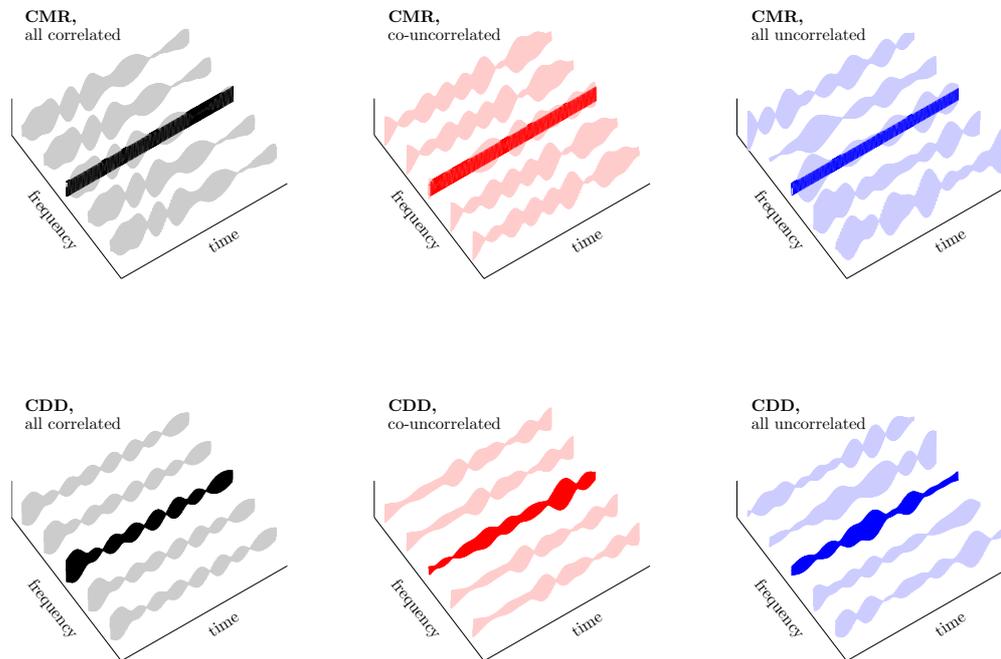


Figure 1.1: Sketches of typical stimulus setups for CMR (top row) and CDD (bottom row) experiments. In every panel, the signal which is supposed to be detected is plotted in dark color. For CMR experiments, the signal is a tone, which is indicated by the flat envelope. For CDD experiments, the signal is a narrow band noise with inherent amplitude fluctuations. The maskers, indicated by light colors, are narrow noise bands with inherent amplitude fluctuations. Note that for CMR there is a masker at the signal frequency, while for CDD the signal is spectrally separated from the masker bands. There are three possible correlation conditions: all noise bands can be correlated (left); the maskers away from the signal frequency can be comodulated while the noise band at the signal frequency has independent amplitude fluctuations (middle); or all noise bands can be uncorrelated (right). There is a second kind of CMR experiments which is not included here, in which there is only one masker centered on the signal frequency which may either be used as it is or which may be modulated by multiplication with lowpass noise.

condition in the bottom row of Fig. 1.1). This is generally true if the comodulated maskers and the signal have no correlated amplitude fluctuations. If signal and maskers fluctuate in the same way (the all correlated condition in the bottom row of Fig. 1.1), then thresholds are similar to the case with uncorrelated maskers. The threshold difference in this kind of experiments has been termed *comodulation detection difference* (CDD). After the first psychophysical experiments on CDD by McFadden (1987) and Cohen and Schubert (1987), further experiments with human listeners have been conducted. Many of them are discussed in Borrill and Moore (2002). CDD studies with animals have been performed on a behavioral (Jensen, 2007;

Langemann and Klump, 2007) and a neurophysiological basis (Bee et al., 2007).

The differences between CMR and CDD are subtle: In CMR experiments, a tonal signal is masked by at least one noise band centered on the tone, while in CDD experiments a narrow band noise signal comes along with at least one additional noise band which usually does not spectrally overlap with the signal. This is illustrated in Fig. 1.1. Note that there are two classes of CMR experiments: Either the masker is a single noise band, or the masker consists of several narrow noise bands. Only the second kind of CMR stimuli is sketched in Fig. 1.1. A more detailed overview and comparison of CDD and CMR experiments can be found in Sect. 3.2.

There are several reasons for examining CMR and CDD experiments: First, due to the ubiquitous presence of comodulated sounds in the environment, signal detection in comodulated noise is an everyday task for the auditory system and is therefore of interest as such. Second, understanding the psychophysical and neurophysiological foundations of the CMR and CDD effects can lead to an improved comprehension of signal processing in the auditory system. Specifically, it may allow for finding stimuli which are particularly well suited for the auditory system. With this, one can devise ways of manipulating stimuli in such a way as to support the auditory system in challenging situations, e.g. for hearing-aid users. Third, it is of interest to replicate the excellent capabilities of the auditory system in technical applications, for which understanding specific effects like CMR and CDD can lead to important progress. In order to approach these goals, it is important to not only perform neurophysiological and psychophysical experiments but also to combine insights from these experiments into modeling frameworks which can explain the observed effects on a physiological basis. By additionally abstracting and simplifying models wherever possible, the very basic principles leading to the experimental results can be captured. This is the approach taken in the present thesis.

The present study centers around the topic of modeling auditory signal detection in comodulated noise for birds and mammals. All models described here focus on signal processing steps of the peripheral auditory system up to the level of the auditory nerve. These processing steps are **frequency filtering**, **envelope extraction**, and **compression**. They are comparable for all animal classes under consideration. The human auditory system may serve as an example to illustrate the auditory pathway and the physiological correlates of the main model processing steps. Therefore, the basic layout of the human auditory system will be reviewed briefly in the following. More details can be found e.g. in Moore (2003) and Pickles (1988), being the sources for the figures stated in the following paragraph.

Human listeners can detect sounds in a frequency range of roughly 20 Hz to 20,000 Hz. In this frequency range, reasonable levels of sounds that can be detected lie approximately between 0 dB SPL and 120 dB SPL, where SPL means *sound pressure level* and measures the sound pressure on a logarithmic scale relative to a reference sound pressure of 20 μ Pa. Sound arriving as longitudinal air pressure waves at the ear is initially transduced to vibrations of the tympanic membrane. This sets the three small middle ear bones malleus, incus, and stapes into motion which in turn cause vibrations of the oval window, a membrane covered opening of the cochlea. Vibrations of the cochlear fluid induced by the motion of the oval window elicit vibrations of the basilar membrane, which due to its mechanical properties responds strongest to vibrations of a certain frequency at certain places: at the base (located at the oval window) the basilar membrane resonates with high-frequency sounds, while at the apex (located at the tip of the cochlea) low-frequency sounds generate the strongest response. Therefore, certain locations on the cochlea correspond to certain frequencies in the incoming sound mixture. This corresponds to the **frequency filtering** in the auditory system and may be modeled with a filterbank (e.g. Hohmann, 2002). Deflection of the stereocilia of inner hair cells caused by basilar membrane motion then results in depolarization of the cells. This leads to a release of neurotransmitter into the synaptic cleft and finally to the generation of action potentials in neurons of the auditory nerve. The timing of action potentials is coarsely locked to the phase of the basilar membrane vibrations for low frequency pure tones of up to approximately 4 kHz. For higher frequencies, the phase locking to the stimulus fine structure deteriorates. Regardless of the phase locking to the fine structure the firing rates of auditory nerve neurons roughly follow the **envelope** of a stimulus (see e.g. Joris et al., 2004). The range of audible sound levels (more than 120 dB) covers a range of sound pressures of more than six orders of magnitude. This very broad range is not transformed into an equally broad range of vibration velocities and amplitudes on the basilar membrane. Rather, the range of possible vibration amplitudes is compressed by an active physiological mechanism which is mainly caused by the outer hair cells. This **compressive nonlinearity** is active for sound pressure levels between about 30 dB SPL and 90 dB SPL and means that a change of one dB in the incoming sound pressure results in a change of less than one dB in the basilar membrane vibration amplitude. The roughly 30,000 afferent auditory nerve neurons transmit the action potentials elicited by the roughly 3,500 inner hair cells (Dallos, 1992) to the next stages of the auditory pathway, which among others pass the brainstem and the thalamus and finally reach the auditory cortex.

For the purpose of the present work this very brief summary of auditory

processing stages suffices although it can by no means be complete. The models derived and described here make many simplifying assumptions in order to allow for a view at the essential processes influencing signal detection in comodulated noise. Some of the simplifications made in the course of the present work are that the influence of the outer and middle ear on the basilar membrane vibrations are neglected and that the processing of nerve activity on higher processing stages is reduced to effective descriptions such as temporal averaging and decision making. Still, the interdisciplinary work described here allows for insights into the basis of signal detection in comodulated noise and can hopefully inspire not only further research in the auditory sciences but also in other areas of science and technology where signals need to be detected in comodulated noise.

There have been modeling efforts for CMR and CDD experiments long before the completion of this thesis. From the beginning the CMR effect was thought to be based on comparisons of the outputs of different filters in the auditory system (e.g. Hall et al., 1984): filter outputs with coherent amplitude fluctuations were thought to be grouped together, thus enhancing the possibility to distinguish the masking noise from the tonal signal, while filter outputs with uncorrelated amplitude fluctuations were assumed to be rather perceived as distinct auditory objects, competing with the “signal object” for the attention of the listener. This conclusion was drawn because noise energy added outside the critical bandwidth of the signal centered filter clearly contributed to the release from masking (e.g. Schooneveldt and Moore, 1987). Thus, CMR was mostly thought to arise from across-channel processes, referring to the different filters of the auditory system as channels. Another possible across-channel process explaining CMR may be that the outputs of channels remote from the signal can indicate times during which the masking of the signal is relatively low and therefore the signal may be detected more easily (*listening in the valleys*, e.g. Buus, 1985). Yet another possible explanation for CMR that has been described by Hall et al. (1984) is suppression¹: Masker contributions far from the signal frequency may result in suppression of the masker peaks on the signal and therefore increase the signal-to-noise ratio for small signal amplitudes (Hall et al., 1988). Due to the fact that the influence of suppression should not be symmetric with respect to its origins above or below the signal frequency, suppression could be excluded as the main contribution to the CMR effect because the amount of CMR is mainly symmetric for masker bands above or below the signal frequency (Hall et al.,

¹Suppression describes the fact that excitation or neuronal activity in response to a certain frequency can be reduced due to the presence of other frequency components in a sound (e.g. Moore, 2003).

1984). However, possible contributions of suppression to CMR are discussed in Ernst and Verhey (2006). Other possibilities for understanding the CMR effect qualitatively may be that the auditory system detects changes in the correlations between outputs of different channels (Richards, 1987; van de Par and Kohlrausch, 1998a) or uses an equalization-cancellation mechanism for further processing of channel outputs (Buus, 1985). These and other conceptual explanations for CMR are summarized in Verhey et al. (2003).

Quantitative psychophysical models for CMR based on the output of only one filter have been proposed by (i) Berg (1996), (ii) Verhey et al. (1999), and (iii) Ernst and Verhey (2006). (i) The model by Berg (1996) uses the output of only one very broadly tuned filter with a bandwidth larger than 2 kHz. By means of half-wave rectification and low-pass filtering, a measure similar to the envelope of this wide-band filtered stimulus is calculated. The amplitude spectrum of this quasi-envelope is used for detecting the presence of the signal in a task comparable to CMR. The quantitative predictions of that model fit quite well to corresponding experimental results, but they are not extended to other CMR or CDD situations. (ii) The model by Verhey et al. (1999) is a true within-channel model which holds for a class of CMR experiments with one broadband masker which may be comodulated by multiplying it with a lowpass noise. Its main idea is to use a model of the effective signal processing of the auditory system for modulation detection (Dau et al., 1997) for simulating signal detection thresholds. The output of a single gammatone filter at the signal frequency is half-wave rectified, low-pass filtered, and further processed by several nonlinear feedback loops. This processed stimulus is subsequently analyzed with a modulation filterbank, and signal detection is based on comparison of the filterbank output with a previously estimated template for a stimulus with a supra-threshold signal. It is concluded that signal detection is based on the effective reduction of the modulation depth in a stimulus with signal compared to a masker alone stimulus. (iii) In the study by Ernst and Verhey (2006) the role of suppression for CMR is investigated by within-channel model simulations with the main processing step being a dual-resonance nonlinear filter (DRNL, see Meddis and O'Mard, 2001). After modeling the influence of the outer and middle ear on the stimulus with an appropriate filter, the stimulus is processed with a DRNL filter which mimics the suppressive nonlinear characteristics of the cochlea. After this stage the stimulus is squared and smoothed using a sliding window constructed from three exponential functions modeling forward and backward masking (see also Plack et al., 2002). Signal detection in the model is based on the ratio of the maxima of the output of this processing with and without signal. The simulations with this model can explain main effects

in accompanying diotic CMR measurements, while dichotic² experiments are not well predicted. Differences between model and experiment are attributed to higher level processing steps like wide-band inhibition at the level of the cochlear nucleus (e.g. Pressnitzer et al., 2001; Neuert et al., 2004).

Possible neurophysiological correlates of CMR are reviewed in Verhey et al. (2003). The wide-band inhibition mentioned above can reduce the response of neurons tuned to the signal frequency selectively at times when the comodulated masker bands have a relatively high amplitude and provide little inhibition at times when the masker bands have a low amplitude. This tends to emphasize times with a relatively high signal-to-noise ratio and therefore facilitates signal detection in the case of comodulated maskers (see Pressnitzer et al., 2001; Verhey et al., 2003, for more details). These experimental findings have been successfully implemented in a computational model (Meddis et al., 2002), which can be used to explain psychophysical experiments with the specific set of stimuli used for that study. A second possible neuronal mechanism which may underly the CMR effect is a deterioration of firing rate locking to the masker envelope which may be introduced by the signal tone (Nelken et al., 1999; Las et al., 2005). A computational model which is able to reproduce these experimental observations has been described by Fishbach and May (2003), where essentially the rectified time-derivative of the log-compressed envelope of the stimulus is encoded by neural activity.

Most modeling efforts for comodulation experiments have been put into the CMR effect. The basic ideas of qualitative models for CDD are the same as those for explaining the CMR effect. However, while CMR has been understood as having important across-channel contributions, there have been several suggestions attributing the CDD effect mainly to within-channel effects (Fantini and Moore, 1994; Borrill and Moore, 2002; Moore and Borrill, 2002). Due to the fact that the qualitative explanations for CDD are very similar to those for CMR, they are only briefly mentioned here: Several authors (e.g. McFadden, 1987; Cohen and Schubert, 1987; Hall et al., 2006) discuss the possible role of auditory grouping in CDD experiments. Mechanisms comparable to *listening in the valleys* are discussed e.g. by Cohen and Schubert (1987), Fantini and Moore (1994), and Moore and Borrill (2002). The reduction of the modulation depth of the maskers as a possible cue for the signal is discussed in Wright (1990), who also proposes mechanisms such as equalization/cancellation or correlation detection as possible bases for CDD.

²Presenting a stimulus *diotically* to a listener means supplying both ears with exactly the same sonic waveforms. A *dichotic* presentation implies that both ears are exposed to different sounds (in the present case masking noise in one ear and signal in the other ear).

Fantini and Moore (1994) describe that temporal fine structure cues may give rise to CDD. Another explanation for CDD which was discussed in McFadden (1987) is that combination bands on the cochlea may provide cues for detecting the signal band.

To the best of the author's knowledge, there is only one quantitative model for CDD, and apart from the work presented in the current thesis, no neuronal models for the CDD effect have been described in the literature. Borrill and Moore (2002) and Moore and Borrill (2002) describe the only quantitative psychophysical within-channel model in which it is assumed that the auditory system can detect times during which the signal-to-masker ratio is above a certain criterion. Over the course of the signal presentation, these times are summed to the cumulative detection time which is assumed to be a measure of the signal detectability. This *dip-listening* model can quantitatively explain accompanying CDD experiments with the drawback that it is not clear how the times of high signal-to-masker ratios are detected by the auditory system. In Moore and Borrill (2002), the mechanism leading to a special focus on these times is proposed to be suppression, although no direct experimental evidence for this suggestion is given.

To sum up the modeling efforts on comodulation experiments, it can be said that up to now the main focus of modeling has been laid on the CMR effect. The quantitative psychophysical models have the drawback that their neurophysiological realization is not entirely clear. The models based on neurophysiological findings on the other hand have many adjustable parameters and still need to make assumptions which are not yet experimentally verified. Both kinds of models have not yet been applied to a broad range of CDD *and* CMR stimuli.

The modeling approaches described in the present thesis pose a more general view on comodulation experiments because they are applied to CDD as well as CMR situations. At least one of the psychophysical models presented in this dissertation is directly motivated from neurophysiological experiments and therefore its neuronal feasibility is known. Despite this proximity to neurophysiology, the model has only three free parameters, which makes it easy to analyze and relies on the fewest possible assumptions. This parsimony is also maintained for the second main model detailed in the thesis, such that a broad range of experimental results can be understood with a minimal set of prerequisites.

1.1 OVERVIEW

The present thesis is organized in a cumulative way. The three main chapters 2, 3, and 4 are modified versions of either published or submitted manuscripts and can therefore largely be read independently. However, Chapter 3 describes an enhanced version of the model derived in Chapter 2, thus referring back to it at some points. Chapter 4 describes a self-contained model which takes a step back and looks at the influence of a particular processing step, the compression, on signal detectability. Differences between the published or submitted manuscripts and the chapters in this thesis are summarized in the footnotes at the beginning of each chapter. Due to the modular composition of the present work, several concepts are repeated at different places. The reader is kindly asked to pardon these recurrences and view them as aid for focusing on separate chapters independently.

Chapter 2 consists of two main parts, the first of which has been published in a modified form (Buschermöhle, Feudel, Klump, Bee, and Freund, 2006). The aim of that first part is to model neuronal responses to CDD stimuli and to relate the modeling results to human psychophysics of the CDD effect. To approach this aim, the results of neurophysiological recordings of the activity of field L2 neurons³ of starlings during the playback of typical CDD stimuli are briefly summarized. It is demonstrated that the firing rates of these neurons show different patterns with increasing signal level depending on the correlation condition of the background noise. These differences are explained with a simple model assuming that the mean firing rate of the recorded neurons codes the mean compressed envelope of the filtered stimulus consisting of the noise masker and the signal. Therefore, the observed neuronal activities can be explained by basic peripheral processing steps. The model is formulated in a largely analytically tractable manner such that the reason for the differences between the noise correlation conditions may be identified as interference due to phase correlations. The model predictions for neural activity in response to CDD stimuli are compared to experimental results for an example recording site, and neurophysiological signal detection thresholds obtained from the model are compared qualitatively to psychophysical signal detection thresholds. The model derived in this part of the chapter comprises the first quantitative model for CDD which is based on physiological data.

The second part of Chapter 2 is a modified extract from Bee, Buschermöhle, and Klump (2007). Its main goal is to test the validity of the model introduced in the first part. This is done by numerically simulating the pro-

³Field L2 in birds is comparable to the primary auditory cortex of mammals.

cessing steps of the model for exactly those stimuli that were used during the neurophysiological experiments, accounting for the frequency tuning and rate response characteristics of each individual recording site. Mean neuronal firing rates obtained from these simulations are compared to the actual recordings, and the neurophysiological signal detection thresholds and CDDs obtained with the model are compared to those found experimentally. It is found that the numerically simulated model can explain large parts of the experimental data with relatively few parameters.

Chapter 3, which has been submitted for publication in a slightly varied form (Buschermöhle, Verhey, Feudel, and Freund, 2007b), intends to demonstrate the scope of the model introduced in Chapter 2. The original model for CDD stimuli is expanded to be applicable to CMR stimuli as well. By including a signal detection step, the model can be applied to various psychophysical CDD and CMR experiments from the literature. With only three adjustable parameters, the model which only incorporates peripheral within-channel effects is capable of reproducing main aspects of the respective experiments. This procedure allows for a discussion of the possibilities and limitations of the model. In summary, the chapter describes the first modeling framework which is applied to various CDD and CMR stimuli at the same time and which is based on neurophysiological findings.

Chapter 4 has been submitted for publication (Buschermöhle, Feudel, and Freund, 2007a). Its main intention is to discuss the influence of only one signal processing step, compression, on signal detection in comodulated noise. In order to do this analytically, a simple form of comodulated noise is defined and its general properties are analyzed. Then a variant of the signal-to-noise ratio (SNR) is defined which is in the following used for the analysis of the detectability of signals. This is done in simulations as well as by approximating the power spectrum of compressed stimuli analytically. The final result stating that improved detection of signals in comodulated noise is already possible after applying a compressive nonlinearity to the stimulus is shown to be true for many different forms of modeling compression. Therefore, an estimate of the influence of compression on the CDD and CMR effects may be given, and the modeling results are discussed in the light of the experimental literature. The model discussed in Chapter 4 is conceptually different from the models of the previous chapters, and it is not directly motivated from experiments. Nevertheless it provides a significant contribution to the understanding of comodulation experiments because it can be formulated analytically and it may be applied to CMR *and* CDD stimuli.

The summarizing statements in Chapter 5 briefly review the preceding chapters and explain the broader significance of the present work. Possible future directions of research are described as well.

CHAPTER 2

SIGNAL DETECTION IN COMODULATED NOISE¹

2.1 ABSTRACT

Signal detection in fluctuating background noise is a common problem in diverse fields of research and technology. It has been shown in hearing research that the detection of signals in noise that is correlated in amplitude across the frequency spectrum (comodulated) can be improved compared to uncorrelated background noise. The mechanism leading to this effect is shown to be a general phenomenon which may be utilized in other areas where signal detection in comodulated noise needs to be done with a limited frequency resolution. A model based on neurophysiological experiments is introduced which can explain improved signal detection in comodulated noise. The proposed signal detection scheme evaluates a fluctuating envelope, the statistics of which depend on the correlation structure across the spectrum of the noise. In the model, signal detection does not require a sophisticated neuronal network but can be accomplished through the encoding of the compressed stimulus envelope in the firing rate of neurons in the auditory system. The proposed model is discussed in two versions. First, an analytically tractable model is used to demonstrate basic mechanisms. Then, more detailed numerical simulations are performed allowing for comparisons between model and experimentally obtained firing rates as well as observed and predicted signal detection thresholds.

¹Sects. 2.1–2.4 have been published with modifications (Buschermöhle, Feudel, Klump, Bee, and Freund, 2006). The main differences compared with the published version are: two more subplots in Fig. 2.1, a more detailed description of the experiments in Sect. 2.3, the addition of Fig. 2.3, a second subplot in Fig. 2.4, the insertion of Eq. 2.12, the omission of the conclusions section, and the inclusion of two appendices. Section 2.5 is an extract from the submitted manuscript Bee, Buschermöhle, and Klump (2007). The text represents only the modeling part of the submitted manuscript and has been modified in order to fit into the present chapter. Table 2.1 has two columns less than the corresponding table in the submitted manuscript, and Fig. 2.7 has two rows less than the corresponding figure in the submitted manuscript. All footnotes in this chapter are only included in this dissertation.

2.2 INTRODUCTION

Deciding if a signal is present in background noise is the central problem of signal detection. The statistical properties of noise masking a signal depend on the origin of fluctuations: the standard property of Gaussian white noise is a consequence of the central limit theorem when assuming the fluctuations to arise from a superposition of many independent individual impacts, as in Brownian motion, whereas exponentially correlated red noise results from the Ornstein-Uhlenbeck process (Gardiner, 2004). A not so frequently investigated but nonetheless important situation is the case of *comodulated noise*. It occurs when amplitudes at different regions of the frequency spectrum fluctuate in a coherent, or temporally correlated, fashion. This *comodulation* can be generated when sound pressure waves propagate through turbulent air (Richards and Wiley, 1980) or as a result of the biomechanics of sound production in humans and other animals (Singh and Theunissen, 2003). Random amplitude fluctuations that may be correlated across the frequency spectrum occur also in other natural phenomena, for example when (star-)light or radio waves are influenced by turbulences in the atmosphere (Tatarski, 1961), and in engineering and technology applications, such as radar (Xu et al., 2004) or sonar (Abraham and Lyons, 2004). Hence, the detection of signals in comodulated noise is a problem of general interest across disparate disciplines.

The auditory system continually performs signal detection tasks. One familiar example is speech perception in noisy social environments, such as a cocktail party (Bronkhorst, 2000). Background noise that is hampering the detection of a signal can be called masking noise. We believe a better understanding of signal detection can be achieved by modeling these processes in the vertebrate auditory system, which appears to have evolved to exploit comodulated noise for lowering detection thresholds (Klump, 1996). Here, we propose such a model based on neurophysiological recordings from the auditory forebrain of the European starling (*Sturnus vulgaris*).

In both humans and starlings, thresholds for detecting acoustic signals in noise can be lower in comodulated noise compared to unmodulated noise or incoherently modulated noise with independent amplitude fluctuations across the spectrum (e.g. Verhey et al., 2003). One example of this general phenomenon is the comodulation detection difference (CDD, see McFadden, 1987). Experimental tests of the CDD effect in humans (e.g. McFadden, 1987; Cohen and Schubert, 1987; Moore and Borrill, 2002) and starlings (Langemann et al., 2005; Langemann and Klump, 2007) have shown that the detection of an amplitude modulated narrowband noise signal is improved when the amplitude envelopes of several spectrally separated narrowband noise maskers are correlated with each other (i.e., comodulated) and independent

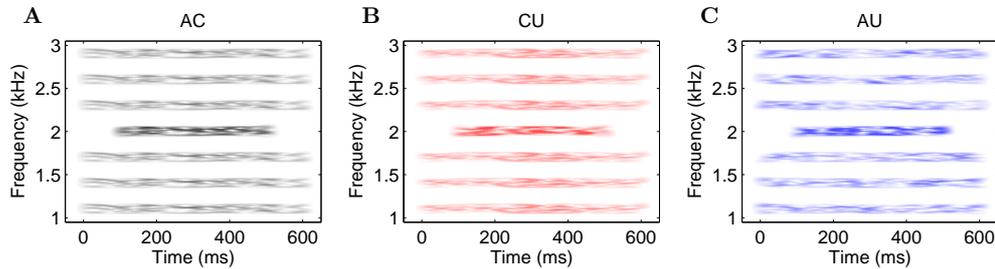


Figure 2.1: Spectrograms of typical stimuli for the (A) all correlated (AC), (B) co-uncorrelated (CU), and (C) all uncorrelated (AU) conditions similar to the ones presented to starlings in the experiments. The signal band here is centered around 2 kHz, flanked by the six masker bands. Depending on the correlation condition, all bands can share a common envelope (i.e. amplitude modulations), the masker bands can have the same envelope while the signal band’s envelope deviates from that, or all bands can have different envelopes (compare also Fig. 3.1). The stimuli shown here are simplified versions of those presented to starlings in electrophysiological experiments (see Sect. 2.4.1).

of that of the signal band (the co-uncorrelated [CU] condition; see Fig. 2.1 B). This improvement in signal detection (i.e., lower detection thresholds) is relative to conditions in which the signal band and each masker band have either a common envelope (the all-correlated [AC] condition; see Fig. 2.1 A) or independent envelopes (the all-uncorrelated [AU] condition; see Fig. 2.1 C). Thresholds in the CU condition can be up to 10 dB lower than those in the AC and AU conditions, which are generally similar (McFadden, 1987; Moore and Borrill, 2002; Langemann et al., 2005).

These threshold differences have been viewed as arising from either within-channel or across-channel effects. This nomenclature refers to the inner ear functioning as a bank of bandpass filters, or channels, in the frequency domain. In the present chapter a simple model is described that uses the output of only one such channel to account for CDD, which is consistent with current psychophysical findings (Moore and Borrill, 2002) and therefore describes CDD as a within-channel effect. The model is based on neurophysiological data recorded in recent CDD experiments with starlings (see Bee et al., 2007) and focuses on a statistical evaluation of the fluctuating envelope of a stimulus comprised of a signal band and a number of masker bands (see Fig. 2.1). The model developed below can predict a number of qualitative and quantitative results from experimental studies of the CDD effect. Moreover, it is based on simple, well-known peripheral processing stages in the auditory system that may be implemented easily into signal detection schemes in other fields. The main model stages are frequency filtering, calculation of the envelope, and compression (i.e. a nonlinear transform of the

envelope).

2.3 EXPERIMENTAL BACKGROUND

The experiments described in this chapter were performed by Mark A. Bee and Georg M. Klump at the Institut für Biologie und Umweltwissenschaften of the University Oldenburg. Electrophysiological recordings were made from six awake and freely moving starlings (three adult females, three adult males) that did not have to perform any specific tasks. Descriptions of the experimental setup and procedure can be found in Langemann et al. (2005) and will be detailed in two future papers (Bee et al., 2007; Langemann and Klump, 2007). The mean firing rates of small populations of neurons (about 5 units) were obtained from extracellular recordings of field L2 (an equivalent of the mammalian primary auditory cortex) during repeated presentations of stimuli similar to those shown in Fig. 2.1. The stimuli consisted of a single signal band centered at the signal frequency and six masker bands which were placed symmetrically around the signal band in the frequency domain. The central frequency of the signal band was set to the frequency at which the neurons were maximally sensitive. The central frequencies of the adjacent noise bands (masker bands) were separated by 300 Hz, and each noise band had a bandwidth of 100 Hz. Each single noise band was generated by multiplying Gaussian lowpass noise with a cutoff frequency of 50 Hz with a sine wave at the desired center frequency of the band. The final masker consisting of six noise bands was generated by adding up the individual noise bands. The level of the masker bands was held constant across all conditions. Overall masker levels of 42.8 dB SPL (sound pressure level) and 77.8 dB SPL were tested. For each of the possible stimulus configurations, 30 random stimuli were generated and played back with a 700 ms silent period between two stimuli. The first 20 artifact free recordings were used to determine the average firing rate of the neurons during the time of signal presentation, taking into account the delay between stimulus playback and onset of neuronal activity (on the order of 15 ms). As the level of the signal band was increased in 5 dB steps, the time- and ensemble-averaged firing rates showed different patterns in the three correlation conditions (Fig. 2.2 A). Two important trends in the data are worth noting. First, neural responses in the AU-condition exhibited higher mean firing rates at low signal levels compared to the other two conditions. Second, the firing rates in the AC-condition exhibited a clearly nonmonotonic pattern. These general findings were observed for all birds, two different signal durations (either 60 ms or 400 ms) and two different signal offsets (either 0 ms (i.e. maskers and signal starting at the same time)

or 100 ms (i.e. signal starting 100 ms later than the maskers)). The reason for studying different signal offsets was that effects of auditory grouping (see Bregman, 1999) were supposed to be analyzed in the experiments. Several sounds may be perceptually grouped into one auditory object if they start at the same time, while they may be perceived as distinct objects if they start at different instances in time. (cf. Bee et al., 2007, for further information). The influence of signal duration and signal delay are of lesser importance for the general model results and are therefore first neglected. They will be discussed in Sect. 2.5.

Signal detection thresholds for the physiological data were determined by calculating the discriminability index d' , which is defined as the difference of the mean spike rates when the signal is present and when it is absent divided by their common standard deviation, assuming both standard deviations are equal (Green and Swets, 1966). Our experimental results showed that the firing rate standard deviations do not depend strongly on signal level (see Appendix 2.7). For the reasonable empirical criterion $d' = 1.8$ (see e.g. Green and Swets, 1966; Hofer and Klump, 2003) the signal is said to have been detected (Fig. 2.2 B). The rank order of detection thresholds ($CU < AC \approx AU$) is consistent with perceptual studies of humans (McFadden, 1987; Moore and Borrill, 2002) and starlings (Langemann and Klump, 2007).

2.4 ANALYTICALLY TRACTABLE MODEL

2.4.1 Model description

The firing rate of auditory neurons can encode the envelope of sound stimuli (e.g. Schreiner and Urbas, 1988; Joris et al., 2004). Therefore, we assume that the time- and ensemble-averaged firing rates observed in our physiological recordings are a measure of the mean stimulus envelope during signal presentation. We first devise an analytically tractable model which translates the mean envelope of a stimulus into a neuronal firing rate. The stimuli that serve as the input to our model are comprised of a number of narrowband noises like the ones shown in Fig. 2.1. The first step of the model consists of spectrally filtering the stimulus by decreasing the amplitude of the masker bands with increasing distance from the signal band. This filtering mimics the frequency selectivity of the inner ear. The filtered stimulus can be understood as the real part of the analytic signal (Gabor, 1946)

$$s(t) = \sum_{k=-M}^M a_k \sum_{n=-N}^N e^{i[(\omega_0+k\Delta\Omega+n\Delta\omega)t+\phi_{k,n}]}. \quad (2.1)$$

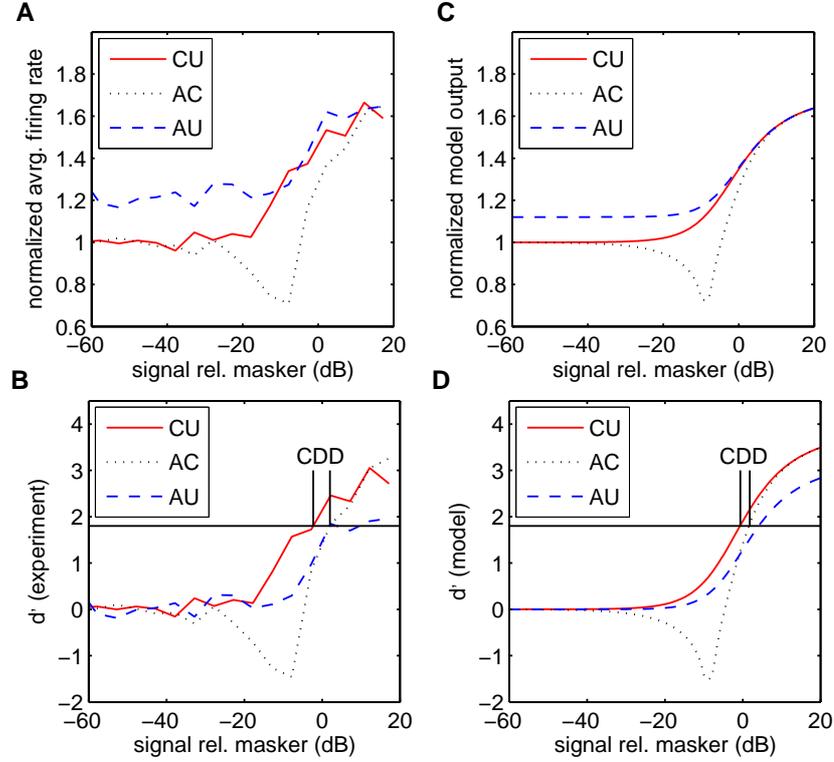


Figure 2.2: **A:** Spike discharge rates from a typical recording site averaged across time and trials and normalized to the average rate if maskers alone are present in the AC condition (recording site: OL03_44_300904, 60 ms signal duration, 0 ms signal offset). **B:** d' -analysis of spike discharge rates. **C:** Expectation values of the time- and ensemble-averaged compressed envelope amplitudes relative to their value if maskers alone are present in the AC condition. A saturating function was applied to account for saturation of the neural firing rate. **D:** d' -analysis of compressed saturating envelopes. For the model, a constant standard deviation $\sigma = 0.18$ in normalized units was assumed. The solid horizontal line in graphs **B** and **D** indicates the detection criterion of $d' = 1.8$. (Model parameters: $\alpha = 0.35$, $a_1 = a_{-1} = 0.8$, $a_2 = a_{-2} = 0.5$, $a_3 = a_{-3} = 0.25$.)

Here, $2M$ is the number of masker bands, $2N + 1$ is the number of sinusoids used to generate the noise bands, ω_0 is the central frequency of the signal, and $\Delta\Omega$ and $\Delta\omega$ define the spacing between noise bands and component sinusoids, respectively (Fig. 2.4 A). The real factors a_k denote the filtered amplitudes of the different noise bands and the signal amplitude is determined by a_0 . Frequency selectivity is modeled by decreasing a_k for increasing $|k|$. For the simple version of the model, we use a symmetric filter ($a_{-k} = a_k$) with a realistic bandwidth for the starling auditory system (Buus et al., 1995, see also Fig. 2.4 B). The phases $\phi_{k,n}$ are the distinguishing element for the different stimulus conditions: if all $\phi_{k,n}$ are randomly chosen from $[0, 2\pi[$,

then an AU-stimulus is generated; if $\phi_{k,n}$ are random for all n but equal for all k , then an AC stimulus is generated; CU stimuli are similar to AC stimuli except for the fact that only for $k, k' \neq 0$ the phases $\phi_{k,n}$ and $\phi_{k',n}$ are the same.

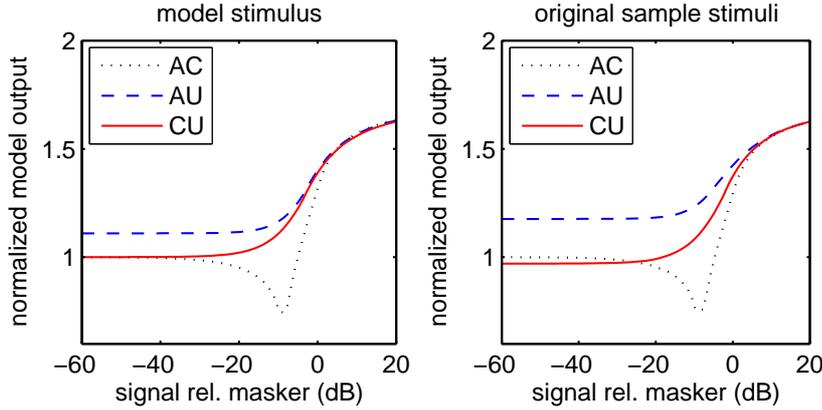


Figure 2.3: Comparison of model results when using the abstract version of the stimuli as described in this chapter (left) and when using original stimuli as they were played back to starlings in the neurophysiological experiments (right). The masker bands of the original stimuli were attenuated in the same way as those of the model stimuli. Then the envelopes were calculated, compressed and finally averaged. The fact that the AC and CU curves for the original stimuli do not start at the same value is due to the limited number of sample stimulus realizations, while the model uses the statistics of all possible realizations. (Same parameters as in Fig. 2.2.)

This is a simplified description of the stimuli for use in the model. The original stimuli were generated by multiplying lowpass noise with pure sinusoids centered at the desired center frequencies of the noise bands (see Sect. 2.3). For the purpose of explaining the general experimental findings, we can use the simplified stimuli. A comparison of model results for the simplified and the original stimuli can be seen in Fig. 2.3. The general results are clearly the same.

The envelope of the filtered stimulus can be easily calculated as the absolute value of the analytical signal $s(t)$. For assessing the envelope statistics, first of all the ensemble average $\langle |s(t)|^2 \rangle_\phi$ can be determined. Using trigonometric addition formulas, one can transform Eq. 2.1 to

$$|s(t)|^2 = \sum_{k,k'=-M}^M a_k a_{k'} \sum_{n,n'=-N}^N \cos(((k-k')\Delta\Omega + (n-n')\Delta\omega)t + \phi_{k,n} - \phi_{k',n'}). \quad (2.2)$$

Making use of further addition theorems and keeping in mind that

$$\langle \cos(\phi_{k,n}) \sin(\phi_{k',n'}) \rangle_\phi = \langle \sin(\phi_{k,n}) \cos(\phi_{k',n'}) \rangle_\phi = 0, \quad (2.3)$$

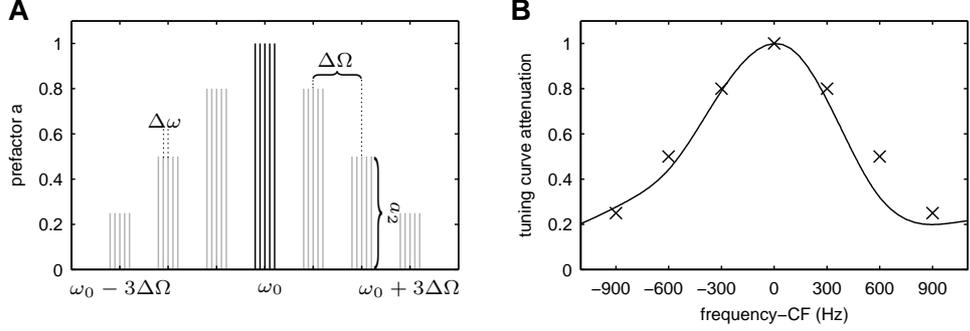


Figure 2.4: **A:** Sketch of the simplified stimulus setup for $M = 3$, $N = 2$, and symmetric $a_k = a_{-k}$. The signal band is plotted darker than the masker bands. Frequency selectivity of the auditory system is accounted for by different amplitudes of the noise bands. **B:** The solid line shows the filter form of an exemplary recording site (OL10_37_061103) normalized to the maximal firing rate elicited by a tone at the CF (characteristic frequency). The prefactors a_k used in **A** are indicated by crossmarks. (Same a_k as in Fig. 2.2.)

$$\langle \cos(\phi_{k,n}) \cos(\phi_{k,n'}) \rangle_\phi = 0 \text{ if } n \neq n', \quad (2.4)$$

$$\langle \cos(\phi_{k,n}) \cos(\phi_{k,n}) \rangle_\phi = \frac{1}{2}, \quad (2.5)$$

one arrives at

$$\langle |s(t)|^2 \rangle_\phi = (2N + 1) \sum_{k,k'=-M}^M a_k a_{k'} \delta(\phi_{k,n}, \phi_{k',n}) \cos((k - k')\Delta\Omega t). \quad (2.6)$$

The δ -terms are one if $\phi_{k,n} = \phi_{k',n}$ and zero if $\phi_{k,n} \neq \phi_{k',n}$. Note that by taking the mean over all possible phases ϕ , the dependence on $\Delta\omega$ disappears while $|s(t)|$ is already independent of the central frequency ω_0 . For large N , the distribution of squared envelope amplitudes at a certain time t can be approximated as an exponential distribution (see e.g. Lawson and Uhlenbeck, 1950). Assuming symmetrical filtering ($a_k = a_{-k}$, see Fig. 2.4 B for a comparison of model filter factors and a filter shape derived from an experimental tuning curve), the mean of this distribution for the three correlation conditions at time t is given by

$$\mu_{\text{AU}}^2 = \langle |s(t)|^2 \rangle_{\phi, \text{AU}} = (2N + 1) \left[a_0^2 + 2 \sum_{k=1}^M a_k^2 \right], \quad (2.7)$$

$$\mu_{\text{AC}}^2(t) = \langle |s(t)|^2 \rangle_{\phi, \text{AC}} = (2N + 1) \left[a_0 + 2 \sum_{k=1}^M a_k \cos(k\Delta\Omega t) \right]^2, \quad (2.8)$$

$$\mu_{\text{CU}}^2(t) = \langle |s(t)|^2 \rangle_{\phi, \text{CU}} = (2N + 1) \left[a_0^2 + 4 \left(\sum_{k=1}^M a_k \cos(k\Delta\Omega t) \right)^2 \right]. \quad (2.9)$$

Here, the correlations of the phases were used to substitute the δ -terms. The cosine terms in Eqns. 2.7–2.9 comprise the important differences between the three correlation conditions. They may be interpreted as interference terms due to correlated phases.

If the distribution of squared envelope values is an exponential distribution with parameter $\mu^2(t)$, then the distribution of envelope amplitudes Y at time t is given by a Rayleigh distribution² with mean $\frac{\sqrt{\pi}}{2}\mu(t)$:

$$p_Y(y, t) = \frac{2y}{\mu^2(t)} \exp\left(-\frac{y^2}{\mu^2(t)}\right) \quad (2.10)$$

Making use of separate time scales (fast fluctuations with multiples of $\Delta\Omega$ and slow fluctuations with multiples of $\Delta\omega$) and a stimulus duration that is much larger than $2\pi/\Delta\Omega$, the final envelope distribution and its moments can be computed by eliminating time. This is done by integrating the time dependent Rayleigh distribution over one period of its parameter $\mu^2(t)$. With Y being the random variable for the envelope amplitude, one gets

$$\begin{aligned} E(Y^\alpha) &= \int_0^\infty y^\alpha \langle p_Y(y, t) \rangle_T dy \\ &= \Gamma\left(\frac{\alpha + 2}{2}\right) \frac{1}{T} \int_0^T |\mu(t)|^\alpha dt \end{aligned} \quad (2.11)$$

as the expectation value of Y^α , where $T = \frac{2\pi}{\Delta\Omega}$ denotes the duration of one period of $\mu^2(t)$. This expression holds for any $\alpha > 0$ and can be used to compute all moments of the distribution of Y^α .

A compressive nonlinearity that is already present in the inner ear (e.g. Köppl and Yates, 1999; Rhode and Recio, 2000; Robles and Ruggero, 2001) is implemented in our model by taking the envelope of the filtered stimulus to the power of α with $\alpha < 1$ (see Eq. 2.11). We use a compression of $\alpha = 0.35$, which is consistent with our own neural data and recordings from the avian auditory nerve (Köppl and Yates, 1999). Mean and standard deviation (SD) of the envelope distributions for this compression are shown in Fig. 2.5.

²A short note on notation: Y denotes the random variable for the envelope amplitudes, while $y \in [0; \infty[$ is a real number expressing possible values for Y .

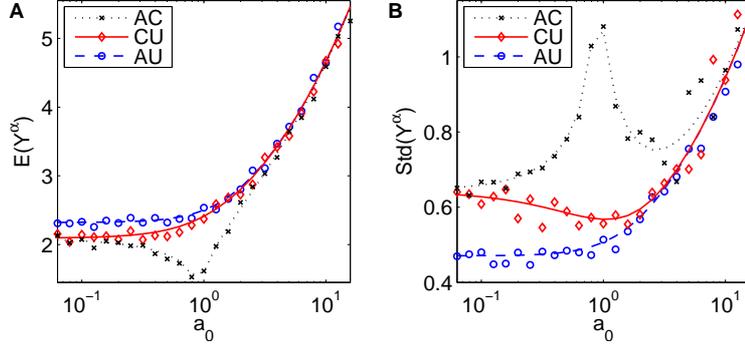


Figure 2.5: Mean (A) and standard deviation (B) of envelope distributions for increasing signal level and constant masker level as predicted from Eq. 2.11 (lines) and as found in sample stimuli (markers) for $M = 3$, $N = 50$, $\Delta\Omega/2\pi = 300$ Hz, $\omega_0/2\pi = 2000$ Hz, and $\Delta\omega/2\pi = 1$ Hz. A compressive nonlinearity is accounted for by the exponent $\alpha = 0.35$. Variability in the data for sample stimuli results from a finite stimulus duration (400 ms). Same a_k as in Fig. 2.2.

In our model, the translation of compressed envelope values to neuronal firing rates is assumed to be linear up to a certain value. Then saturation at firing rate κ is taken into account by the empirical expression

$$\text{firing rate} \propto \begin{cases} x & \text{for } x < x_0 \\ x_0 + \frac{(\kappa - x_0)(x - x_0)}{\kappa + x - 2x_0} & \text{for } x \geq x_0 \end{cases} \quad (2.12)$$

where x is the compressed envelope amplitude normalized to the average compressed envelope amplitude of the AC masker alone. The average firing rate can be computed as proportional to the time and ensemble expectation value of the compressed saturating envelope of the filtered stimulus. The output of this model for realistic parameters $\kappa = 1.7$ and $x_0 = 1.3$ is shown in Fig. 2.2 C (κ and x_0 are given without units because the firing rates are measured relative to the masker alone rate in the AC case). Saturation is needed for comparison with the experimentally obtained firing rates and can be omitted without altering the general effect of different thresholds for the three correlation conditions. Eq. 2.11 can be used in the case of no saturation.

2.4.2 Model results

To estimate the magnitude of the CDD effect predicted by the model, we implemented the same signal detection scheme that was used to determine physiological thresholds. We determined d' by assuming a firing rate standard deviation σ independent of signal level and stimulus condition. This is consistent with our neural data (see Appendix 2.7). Due to this independence,

our detection scheme reduces to a comparison of firing rates in the conditions with and without signal: a certain increase in firing rate indicates signal detection. As shown in Fig. 2.2 D, the model predicts thresholds for the CU, AC, and AU conditions that are similar to those depicted for the example recording site (Fig. 2.2 B). The rank order of thresholds agrees with previous CDD experiments in humans (McFadden, 1987; Moore and Borrill, 2002) and starlings (Langemann et al., 2005). The amount of CDD as determined from our experiments and our model is less than found in psychophysics. This may be understood using the following reasoning: The firing rate standard deviation of small neural populations is larger than that of large populations of neurons. A smaller standard deviation for the calculation of d' results in a vertical expansion of the curves in Fig. 2.2 D and therefore yields an increase in CDD if the detection criterion of $d' = 1.8$ is kept constant.

The threshold differences between the three correlation conditions can be understood by noting two important differences in the mean compressed envelopes (Figs. 2.2 C and 2.5 A). First, there is a prominent “dip” in the AC condition which is present for $0 < \alpha < 2$ (see Appendix 2.8). This dip results from the beat phenomenon and occurs when the overall signal and masker levels are approximately equal. At this signal level, beating due to correlated phases has the biggest effect: destructive interference between the correlated noise bands leads to many time instances during which the stimulus envelope is very close to zero, which leads to a relatively small mean compressed envelope value. The dip causes the corresponding AC discriminability index (d') to exceed the detection criterion at a higher signal level compared to the CU condition (Fig. 2.2 D). Second, the mean envelope of a stimulus comprised only of the maskers ($a_0 = 0$) is higher in the AU condition than in the other two conditions (Fig. 2.2 C). This result is also rooted in the beat phenomenon. In the AU condition, the incoherently modulated masker bands do not interfere strongly with each other, while in the AC and CU conditions interference due to common phases leads to a relative reduction of the average compressed envelope. This explains why d' , which is computed based on the difference between the signal plus masker and masker alone envelopes, crosses the threshold criterion at a higher signal level in the AU condition compared to the CU condition. The influence of the compressive exponent α and the filter bandwidth on these two features of the mean compressed envelopes is discussed in Appendix 2.8. It turns out that the described differences in the mean compressed envelopes are still present with compression $\alpha = 1$ and without frequency filtering. Therefore, the extraction of the mean envelope is the essential step for obtaining threshold differences between the correlation conditions.

2.5 ENHANCED MODEL WITH NUMERICAL SIMULATIONS

In the previous section, we presented a largely analytically tractable within-channel model to explain the CDD effect. The analytical expressions derived there are valid for infinitely long stimuli and idealized frequency filtering and stimulus generation. In this section, we make the model more realistic by numerically simulating the auditory processing stages of the model and using the exact same stimuli as those used in the original experiments. In these simulations, all stimulus conditions that were tested in the experiments were also tested. This means that we looked at two different masker spectrum levels³ (15 dB SPL and 50 dB SPL, corresponding to overall levels of 42.8 dB SPL and 77.8 dB SPL, respectively), two different signal durations (60 ms and 400 ms), and two different signal onset delays relative to the masker (0 ms or 100 ms). For each of the two masker spectrum levels, the neuronal firing rates of 16 recording sites were modeled at the four combinations of signal duration and signal delay.

2.5.1 Model processing steps

Signal processing was performed numerically according to the scheme shown in Fig. 2.6. The input to the model was 20 stimulus realizations with the appropriate levels and combinations of the manipulated variables (Fig. 2.6 A). For each recording site, the stimuli were filtered according to the appropriate frequency tuning curve (Fig. 2.6 B). The frequency tuning curve was smoothed using cubic splines and then a rate response curve as a function of frequency was calculated by averaging over all levels above 42 dB. This rate response curve was normalized to 1.0 at its maximum. The filter attenuation factors across the bandwidths of each noise band in the stimulus were determined from this normalized frequency-rate curve (see Fig. 2.4 B for an example filter shape deduced from a tuning curve). For simplicity, we assumed filter shapes to be independent of masker and signal levels (see Gleich, 1994). The Hilbert envelopes of the filtered stimuli were then computed (Fig. 2.6 C) and compressed (Fig. 2.6 D) by raising them to the exponent $\alpha \leq 1$ (see below). The conversion from compressed envelope $y(t)$ to continuous, time-dependent firing rates was done using the following empirical

³The spectrum level measures the level of a sound within a 1 Hz wide band. The spectrum level L_{spect} is related to the overall level L_{ovrl} by the expression $L_{\text{spect}} = L_{\text{ovrl}} - 10 \log_{10}(BW)$, where BW denotes the bandwidth in Hz (see e.g. Hartmann, 1998).

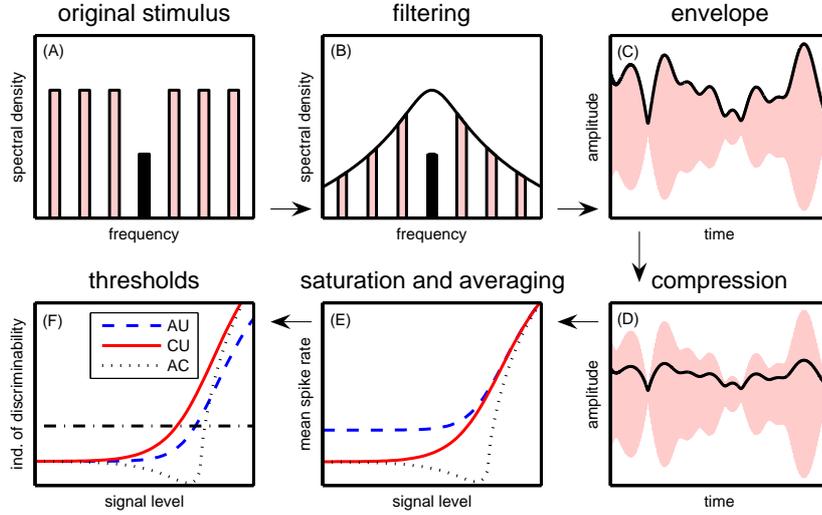


Figure 2.6: Schematic diagram showing various stages of the peripheral model. Going clockwise from top left to bottom left, the subplots show (A) the stylized spectrum of the raw stimulus showing the signal band in black and the flanking bands in gray, (B) the spectrum of the filtered stimulus along with the corresponding filter shape, (C) the time signal with its envelope, (D) the time signal with the compressed envelope, (E) the mean firing rates depending on signal level calculated from the compressed envelopes by using a saturating function (different line styles for the three correlation conditions), and (F) the d' -curves with the signal detection criterion shown by the dash-dotted line.

saturating function (Fig. 2.6 E):

$$r(y(t)) = r_0 + (r_{\max} - r_0) \left(1 - \exp\left(-\frac{y(t)}{c}\right) \right), \quad (2.13)$$

where r is the neuronal firing rate (in impulses/s). The parameters r_0 , r_{\max} , and c are, respectively, the recording site's spontaneous firing rate, its saturating firing rate, and a parameter determining the level at which the recording site reached half of its dynamic range. Finally, the continuous, time-dependent firing rates $r(t)$ were averaged across the signal duration at the appropriate signal onset delay and across the 20 stimulus realizations that were input into the model.

A total of four parameters could be adjusted for each recording site and each combination of signal onset delay and signal duration. Because of the phasic-tonic response properties of field L2 neurons (e.g. Nieder and Klump, 1999), different fitted parameters for r_0 , r_{\max} , and c were required to model responses to the two signal onset delays (0 ms and 100 ms) and the two signal durations (60 ms and 400 ms). Model parameters were determined as follows. The four free parameters (α , r_0 , r_{\max} , and c) were set to realistic initial values

for each recording site and then varied within physiologically realistic boundaries in order to minimize the mean squared differences between the model firing rates and the experimentally obtained firing rates from field L2 neurons. This procedure was performed independently for all four combinations of signal duration and signal onset delay to derive initial estimates of the optimal model parameters. We assumed that the compression α was constant for a given recording site; therefore, we averaged the derived values of the compression α across the four combinations of signal onset delay and signal duration. With α fixed, the remaining three parameters were readjusted for each combination of signal onset delay and signal duration by again minimizing the mean squared differences between the model and experimental firing rates. Hence, the same value of α was used for a given recording site to model the rate-level curves for all 12 stimulus conditions tested at that recording site (3 correlations \times 2 signal onset delays \times 2 signal durations), whereas different values of r_0 , r_{\max} , and c were used to model the rate-level curves for the different combinations of signal onset delay and signal duration.

Finally, the average firing rates were converted to the discriminability index (d') to determine the model's signal detection threshold as the signal level at which d' first exceeded a threshold criterion of 1.8 (Fig. 2.6 F). This is the same threshold criterion used for determining neural response thresholds in Bee et al. (2007). In the model, d' was calculated from the model rate-level curves by subtracting the rate response to the masker alone from the response to the signal plus masker and then dividing by an average standard deviation σ of neural firing rates (see Appendix 2.7). Averaged across experimental conditions, the standard deviations of the impulse rates elicited in responses to the two signal durations were different (mean \pm 95% confidence intervals: 60 ms condition, 50.8 ± 0.4 impulses/s; 400 ms condition, 25.1 ± 0.3 impulses/s); therefore, we used constant standard deviations of $\sigma_{60} = 50.8$ impulses/s and $\sigma_{400} = 25.1$ impulses/s for all recordings with 60 ms and 400 ms signal duration, respectively. These values are the averages of all firing rate standard deviations for the recording sites used in this study obtained experimentally for the respective signal durations. We did not vary the standard deviation based on other manipulated variables (e.g. correlation and signal onset delay); therefore, differences in firing rate variability could not account for any CDD effect in the model results.

2.5.2 Results of model simulations

The rate-level (and d' -level) functions based on the time and trial averaged firing rates output by the model were similar to those observed for the actual

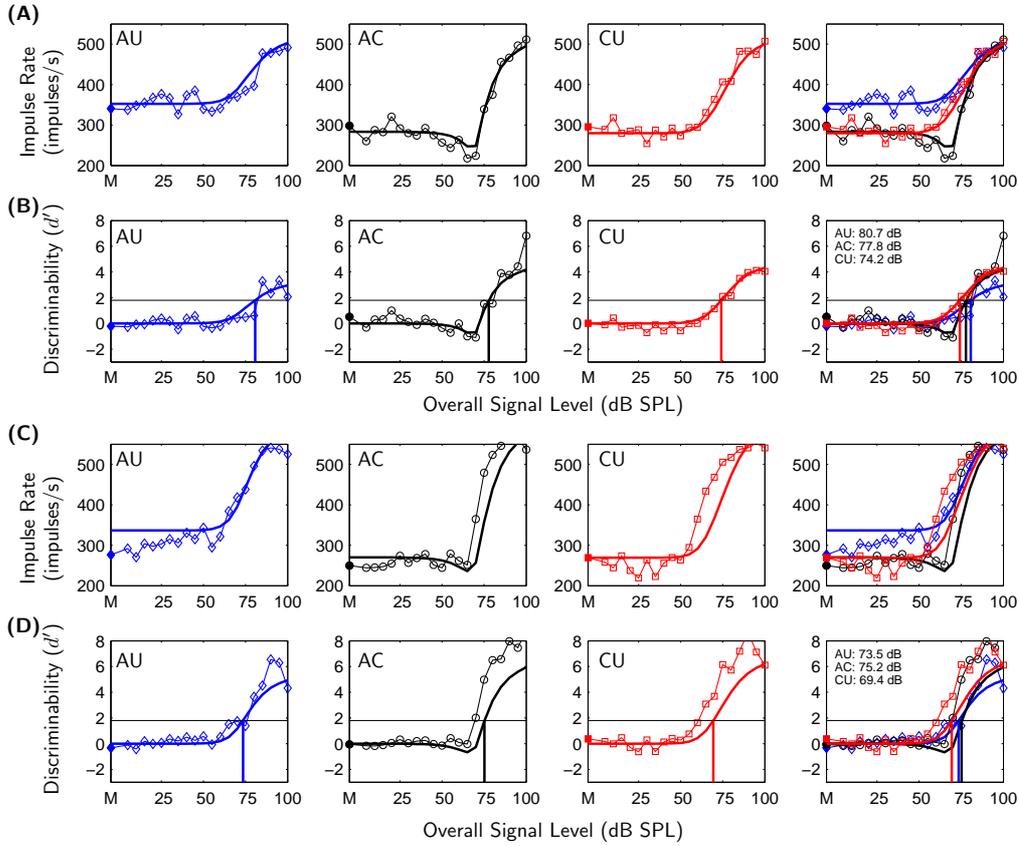


Figure 2.7: Representative rate-level and d' -level functions for neural data and model output. **A:** Rate-level function for a representative recording site (OL03_46_011104) tested with the 50-dB masker spectrum level illustrating the changes in impulse rate that occurred as a function of overall signal level for the AU, AC, and CU correlation conditions (signal onset delay: 0 ms; signal duration: 60 ms). 'M' and the filled data points depict the masker-driven responses in each condition. The plot on the far right depicts data for all three correlation conditions together. **B:** Data from the same recording site and conditions depicted in **A** after converting impulse rates to the discriminability index (d') using signal detection theory (Green and Swets, 1966). Black horizontal lines depict a threshold criterion of $d' = 1.8$. Smooth curves represent the model output. Vertical lines show the point where the d' -level function crossed the threshold criterion. These thresholds are provided in the far right plot. **C** and **D** show the rate-level functions and the d' -level functions from the model output for a recording site with one of the poorest model fits (OL44_32_021104).

rate-level functions of field L2 neurons (Fig. 2.7). While there was some variation in how well the model reproduced the rate-level functions from different recording sites (cf. Figs. 2.7 A, B and 2.7 C, D), the model generally captured two important aspects of real rate-level functions. First, the masker driven responses in the AU conditions were higher than those in the AC and

Table 2.1: The magnitude of the model CDD effect as a function of various stimulus properties and comparisons to physiologically determined CDD effects in field L2 neurons (Bee et al., 2007).

Stimulus Properties			CDD Effect			
Masker Spectrum Level (dB)	Signal Delay (ms)	Signal Duration (ms)	Model CDD (dB)		Neural CDD (dB)	Difference (dB)
			Mean	SD	Mean	
15	0	60	-1.4	1.9	-3.5	2.1
		400	-1.5	1.1	-6.2	4.7
	100	60	-1.0	0.6	-2.8	1.8
		400	-1.7	1.0	-2.3	0.6
50	0	60	-2.1	4.3	-2.7	0.6
		400	-3.9	5.4	-8.3	4.4
	100	60	-2.0	6.4	-4.2	2.2
		400	-3.7	8.6	-7.4	3.7

CU conditions (e.g. Fig. 2.7 A). Second, the model was able to successfully reproduce the prominent “dip” in the rate-level functions of the AC conditions (e.g. Fig. 2.7 A).

Across the eight combinations of signal onset delay, signal duration, and masker spectrum level, the differences between signal detection thresholds generated by the model in the CU and AC conditions were always less than zero, indicating that the model generated a CDD effect in all stimulus conditions tested (Table 2.1; Fig. 2.8 A). The mean magnitudes of CDD in the model’s output across stimulus conditions (-1.0 dB to -3.9 dB) were uniformly smaller than those measured in field L2 (Table 2.1). This underestimation of the CDD effect by the model is directly related to the model’s overestimation of masked neural thresholds. Across stimulus conditions, the model overestimated neural thresholds by about 3.7 dB (range: 0.3 - 7.2 dB; Fig. 2.8 B). Importantly, however, this overestimation was not random with respect to the three correlation conditions. Rather, the model overestimated the mean thresholds in the CU, AC, and AU conditions by 5.8 dB, 3.2 dB, and 1.9 dB, respectively. Given that the actual thresholds in these three conditions were ranked in the opposite order (i.e., $AU > AC > CU$), the threshold overestimations had the effect of reducing the threshold differences among the correlation conditions (and hence the magnitude of CDD) in the model’s output.

The following general trends were similar in both the model and physiological results (Fig. 2.8 B): (i) thresholds were highest in the AU condition,

lowest in the CU condition, and intermediate in the AC condition; (*ii*) thresholds were lower with the 400 ms signal compared to the 60 ms signal; and (*iii*) thresholds were lower at the 50 dB spectrum level compared to the 15 dB spectrum level.

2.5.3 Discussion

In this section, we applied the within-channel model of Sect. 2.4 to a much larger number of recording sites and stimulus conditions than was done there. For most recording sites (e.g. Fig. 2.7 A), the model successfully reproduced the higher masker driven responses in the AU conditions and the prominent “dip” in the AC rate-level functions. Importantly, both of these features are already clearly present in the mean compressed envelope values of the filtered CDD stimuli (see Sect. 2.4). The average compressed envelope of the uncorrelated flanking bands in the AU condition is larger than that of the comodulated flanking bands in the AC and CU conditions due to destructive interference between the correlated bands in the AC and CU conditions. No such destructive interference occurs between the uncorrelated flanking bands in the AU conditions. Destructive interference is also responsible for the dip in the AC rate-level curves. The existing destructive interference between the comodulated flanking bands in the AC and CU correlation conditions is enhanced in the AC condition with the addition of a correlated signal band. Consequently, when the level of the signal approaches the overall level of the flanking bands, the average compressed envelope gets smaller before increasing again once the overall envelope becomes dominated by the signal envelope. No such additional interference occurs when the signal is added in the CU conditions because the signal is not correlated with the flanking bands.

In general, the model also performed fairly well in reproducing the effects of signal onset delay, signal duration, and masker spectrum level on the magnitudes of CDD (Fig. 2.8 A) and masked detection thresholds (Fig. 2.8 B). Two notable and related features of the model’s performance in these simulations are that it underestimated the mean magnitudes of the neural CDD effect and overestimated the masked detection thresholds (Fig. 2.8). There are two general classes of explanation for this shortcoming. First, our model does not include across-channel processing, which some have hypothesized could contribute to the CDD effect (e.g. Cohen and Schubert, 1987; McFadden, 1987). While the model itself may not exclude the possible operation of across-channel processes, little evidence for auditory grouping indicating across-channel processing was found in the patterns of neural thresholds in

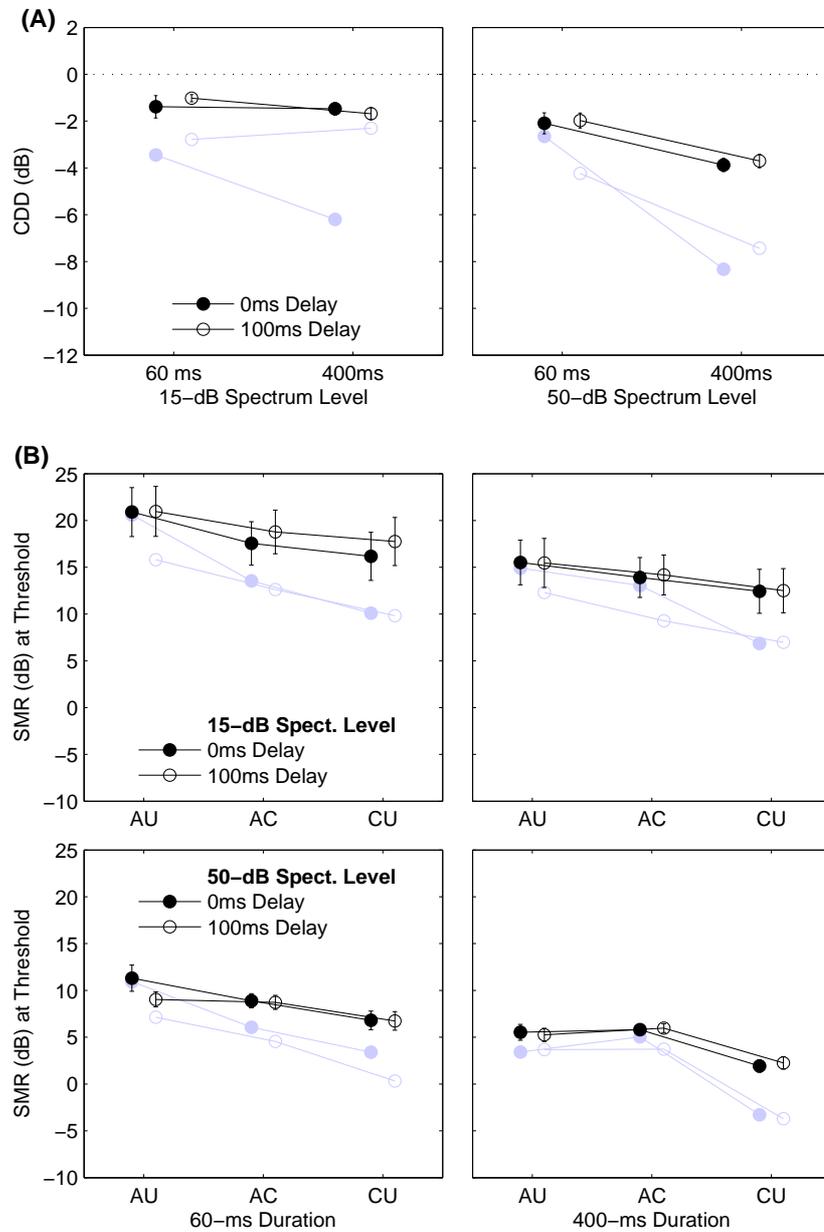


Figure 2.8: CDD and masked detection thresholds from the within-channel peripheral model. **A:** The mean (\pm standard error) magnitude of CDD for the 15 dB (left) and 50 dB (right) masker spectrum levels depicted as a function of signal duration (60 ms or 400 ms) with signal onset delay as the parameter. Negative values indicate a CDD effect. **B:** The mean (\pm standard error) signal-to-masker ratios (SMR in dB) at thresholds for the 15 dB (top) and 50 dB (bottom) masker spectrum levels depicted as a function of correlation condition (AU, AC, and CU) for each signal duration (60 ms or 400 ms) with signal onset delay (0 ms or 100 ms) as the parameter. For comparison to neural data, the experimental results from Bee et al. (2007) are reproduced here in lighter color.

Bee et al. (2007) and in behavioral thresholds in Langemann and Klump (2007). Second, and more likely, the model in its current form does not capture the full breadth of the within-channel processes that may contribute to CDD. Langemann and Klump (2007) discuss these additional within-channel cues in more detail. We believe one likely within-channel mechanism that is not currently implemented in the model, but that could also contribute to CDD, is the suppression⁴ that results from cochlear mechanics and that has been observed in the auditory nerve (see Moore and Borrill, 2002, for further discussion). Suppressive effects due to nonlinear filtering have been modeled using a dual resonance nonlinear filter (DRNL, see Meddis and O'Mard, 2001). Such a signal processing step instead of linear filtering would allow for investigating some influences of suppression on the model. It has not been used in the present work in order to use the measured tuning characteristics of the original recording sites.

2.6 CONCLUSIONS

The present chapter describes a simple model for the CDD effect: general features of rate-level functions obtained from neurophysiological recordings in starlings during playback of typical CDD stimuli can be understood by considering the peripheral signal processing steps filtering, envelope extraction, compression, and averaging. This within-channel processing can even explain many of the detailed features of the neurophysiological observations and therefore argues for understanding CDD as a within-channel effect. Differences in the detailed comparison of model predictions and experimental observations may possibly be explained with the absence of suppression in the model. Apart from neurophysiology, by employing a signal detection scheme based on the index of discriminability d' , the rank order of psychophysical signal detection thresholds in CDD experiments with humans can be reproduced. Using an analytical approach, the differences between correlation conditions can be traced back to interference due to correlated phases. As this kind of interference may also be important in natural noisy situations, the extraction and compression of signal envelopes may be a promising strategy for signal processing in various fields outside the narrow boundaries of hearing research.

⁴See footnote on page 6.

2.7 APPENDIX: FIRING RATE STANDARD DEVIATIONS

For modeling the index of discriminability and therefore signal detection, we assumed the standard deviations of the model firing rates to be independent of signal level, masker level, signal onset, and correlation condition. The only factor influencing the model standard deviations was the signal duration. This model assumption of largely constant standard deviations has been made in order to keep the model as simple and transparent as possible and to avoid additional more complex model assumptions. Also, the firing rate standard deviations cannot be modeled by the envelope standard deviations depicted in Fig. 2.5 B because the firing rate standard deviations measure the variability in spike count within a certain time span, while the envelope standard deviations measure the variability in envelope amplitude for an ensemble of stimuli. Indeed, the CDD effect can already be explained using constant standard deviations, indicating that the firing rate variability is of secondary importance for the CDD effect.

The assumption of constant firing rate variability is a good first approximation to the experimental data. Exemplary traces of firing rate standard deviation σ depending on signal level are plotted for the three correlation conditions in Fig. 2.9 A and D, where results for two typical recording sites are shown at two different masker spectrum levels and signal durations and for the same signal delay. The firing rate standard deviations are determined in the following way: for every one of 20 trials, the temporal mean firing rate is determined as the number of spikes that occur during the presentation of the signal divided by the duration of the signal. σ is then defined as the standard deviation of these 20 mean firing rates.

One can see that each data trace in Fig. 2.9 A and D may be fitted by a straight line which will generally have a slope close to zero. A linear regression for each of the measured data sets of firing rate standard deviation in relation to signal level was performed, and the average slopes of the resulting fits are displayed in Fig. 2.9 B and E. The abscissae in those panels show the different combinations of signal duration and signal delay: The first digit after the ‘s’ may either be a 4 (symbolizing 400 ms signal duration) or a 6 (representing 60 ms signal duration). The digit after the ‘d’ may either be a 0 (meaning 0 ms signal delay) or a 1 (denoting 100 ms signal delay). One can see in the two panels that all slopes are on average in the range of ± 0.15 imp./s/dB which means that the regression lines are essentially flat, given that the average standard deviations are on the order of 20 to 50 impulses per second. This is true for all correlation conditions, masker spectrum levels, and combinations of signal duration and delay. Therefore, one may conclude that the assumption of firing rates being independent of signal level is a

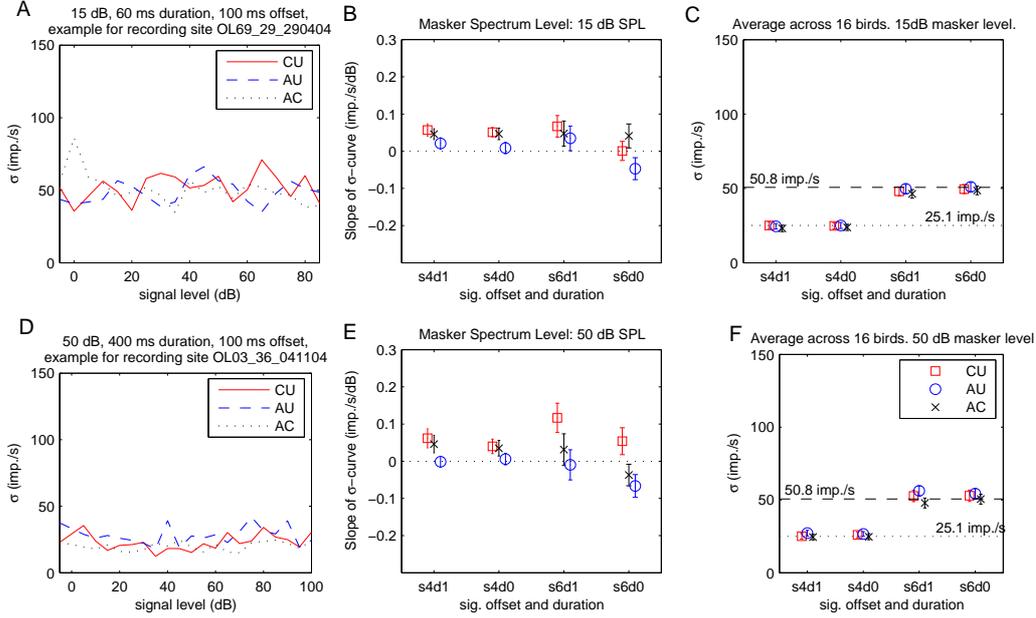


Figure 2.9: Demonstration of constancy of standard deviations. **A** and **D**: Examples for the dependence of firing rate standard deviations σ on signal level for two recording sites. **B** and **E**: A regression line was fitted to each curve of measured standard deviations depending on signal level. The average slopes and their standard errors are plotted. The ticks of the abscissa denote signal duration and signal delay: s4d1 symbolizes 400 ms signal duration and 100 ms signal delay, while s6d0 denotes 60 ms signal duration and 0 ms signal delay. **C** and **F**: Averages and standard errors of firing rate standard deviations across recording sites. Top: 15 dB masker spectrum level, bottom: 50 dB masker spectrum level. The most important difference for the standard deviations is that between 400 ms signal duration and 60 ms signal duration: In the first case, the average standard deviation is 25.1 imp./s, while in the second case it is 50.8 imp./s. These two averages are plotted as dotted and dashed horizontal lines, respectively, in panels **C** and **F**.

reasonable model assumption.

In the above considerations, only the slope but not the absolute value of the firing rate standard deviations has been taken into account. In Fig. 2.9 C and F, the average values of σ are shown. One can see that the main influence on the absolute value of σ stems from the signal duration. For a short signal duration of 60 ms, the firing rates are quite variable with an average standard deviation of 50.8 imp./s. The long signal duration of 400 ms leads to an average standard deviation of 25.1 imp./s.

Although other statistical dependencies may be found in the firing rate standard deviations, the main influence of variability on signal detection has been captured by using two different standard deviations for the different signal durations in the model. Apart from this, it seems reasonable to assume

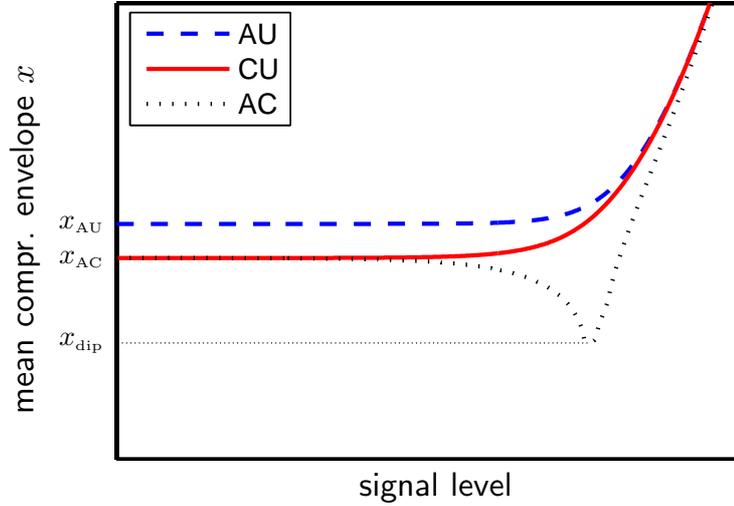


Figure 2.10: Sketch of mean compressed envelope values depending on signal level for typical CDD stimuli. This sketch defines the variables x_{AU} , x_{AC} , and x_{dip} which are used for plotting Fig. 2.11. The mean compressed envelope values for the maskers alone in the AU and AC conditions are denoted by x_{AU} and x_{AC} , respectively. The mean compressed envelope value at the minimum of the dip is denoted by x_{dip} .

constant firing rate standard deviations in order to maintain a simple and transparent model.

2.8 APPENDIX: INFLUENCE OF FILTER BANDWIDTH AND COMPRESSION

The choice of the compressive exponent α and of the bandwidth γ of the auditory filter influence the shape of the modeled rate-level curves and therefore also the differences between these curves for the AC, AU and CU cases. Two features may be used in order to quantify this influence: The reduction in AC masker alone response x_{AC} relative to the AU masker alone response x_{AU} and the minimum of the dip in the AC condition x_{dip} (see Fig. 2.10).

The reduction of x_{AC} relative to x_{AU} caused by the phase correlations can be quantified as $\frac{x_{AC}}{x_{AU}}$. This measure is plotted in Fig. 2.11 A depending on filter bandwidth γ and compressive exponent α . The size of the dip in the AC condition is quantified as the quotient of the dip's minimum and the masker alone response in the AU condition $\frac{x_{dip}}{x_{AU}}$. The deeper the dip, the smaller is this measure, which is plotted in Fig. 2.11 B.

For the compressive exponent, values up to two are considered although only $\alpha \leq 1$ is physiologically realistic. The fact that for any value of $0 < \alpha < 2$ the plotted values in Fig. 2.11 A and B are less than 1 indicates that the

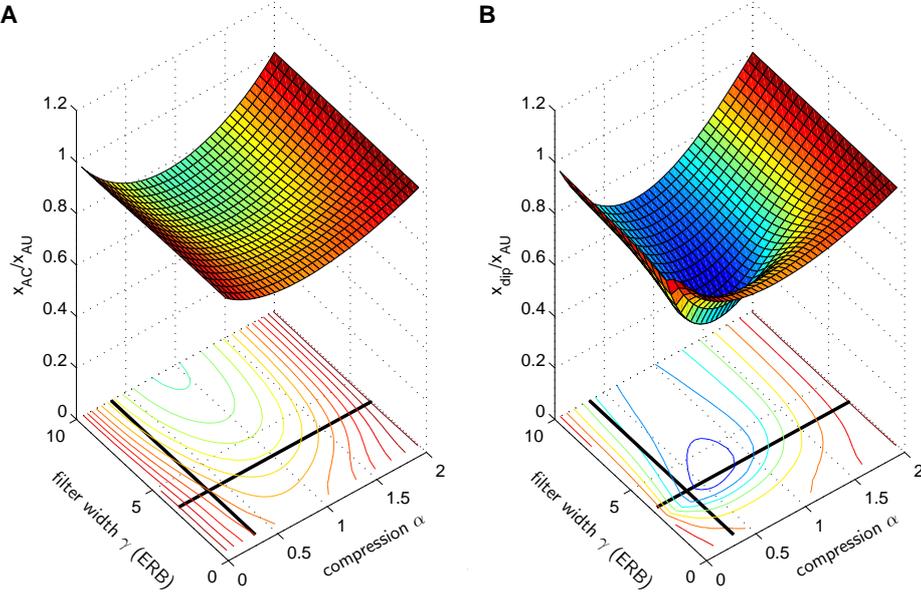


Figure 2.11: Influence of filter bandwidth γ and compression α on the shape of the AC curve. **A:** The relative reduction of masker alone response in the AC case compared to the AU case is quantified by plotting $\frac{x_{AC}}{x_{AU}}$. Panel **B** shows the depth of the dip in the AC condition relative to the masker alone response in the AU condition, i.e. $\frac{x_{dip}}{x_{AU}}$. The variables used here are explained in Fig. 2.10. The filter factors used for this figure are derived from a gammatone filter (Hohmann, 2002) centered on the 2 kHz signal. The filter width γ is measured on a human ERB scale as described in Sect. 3.3.2. The thick solid lines in both subplots mark the parameter values corresponding to those parameters used in Sects. 2.2 to 2.4. The stimulus setup is the same as in Fig. 2.5.

described differences in the model rate-level curves are present even without compression, as stated in Sect. 2.4.2. Also, one nicely sees that for $\alpha = 2$ all curves are the same, which is indicated by the fact that for that parameter value both measures are exactly equal to 1. There is an optimal compression for causing the deepest dip. This compression is $\alpha \approx 0.65$, which is larger than the value of the compressive exponent used in Sects. 2.2 to 2.4. It is also slightly larger than the range of compressive exponents reported in Köppl and Yates (1999) for avian auditory nerve neurons. The existence of an optimal compression can be explained by considering that for $\alpha = 0$ as well as for $\alpha = 2$ all three model curves must be the same⁵, such that for intermediate values of α there must be a minimum in both measures plotted

⁵For $\alpha = 0$, all compressed envelope values are identical to 1; for $\alpha = 2$, the Parseval theorem states that $E(Y^2)$ equals the integrated spectral content of the stimulus, which is exactly the same in the different correlation conditions.

in Fig. 2.11 A and B, given that the phase correlations lead to reduced mean compressed envelopes.

Also the filter bandwidth γ influences the shape of the model curves. Here, we use filter factors a_k obtained from a human gammatone filter centered on the signal frequency and measure the filter bandwidth in ERB as described in Sect. 3.3.2. The filter factors used throughout Sects. 2.2 to 2.4 which have been used for modeling the avian auditory filter can be approximately obtained from the gammatone filter with $\gamma = 3.6$ ERB, which is indicated by the thick solid lines in Fig. 2.11 A and B. One can see that increasing the filter width generally increases the differences between the AU and AC curves, although for the depth of the dip there is an optimal filter width. With a broader filter, the outer masker bands gain more influence on the mean compressed envelope and therefore the phase correlations in the AC condition lead to a more pronounced reduction in the masker alone response. The nonmonotonic dependence of the dip size on γ may be understood by noting that there is a tradeoff between the reduction of mean compressed envelopes due to correlated phases and the growth of mean compressed envelopes due to increased masker and signal power.

CHAPTER 3

A UNIFYING MODEL APPROACH TO COMODULATION DETECTION DIFFERENCE (CDD) AND COMODULATION MASKING RELEASE (CMR)¹

3.1 ABSTRACT

Natural sounds often exhibit correlated amplitude modulations at different frequency regions, so called comodulation. Therefore, the ear might be especially adapted to these kinds of sounds. Two effects have been related to the sensitivity of the auditory system to common modulations across frequency: comodulation detection difference (CDD) and comodulation masking release (CMR). Research on these effects has been done on the psychophysical and on the neurophysiological level in humans and other animals. Until now, models have focused only on one of the effects. In this chapter, a simple model based on data from neuronal recordings obtained during CDD experiments with starlings is discussed. This model is capable of qualitatively reproducing psychophysical signal detection thresholds in response to CDD *and* CMR stimuli. Moreover, it is largely analytically tractable. The model is based on peripheral processing and incorporates the basic steps frequency filtering, envelope extraction, and compression. Signal detection is performed based on changes in the mean compressed envelope of the filtered stimulus. Comparing the results of the model with data from the literature, the scope of this unifying approach to CDD and CMR is discussed.

3.2 INTRODUCTION

One of the prominent aspects of auditory scene analysis is the formation and discrimination of auditory objects (Griffiths and Warren, 2004). Having

¹A modified version of this chapter has been submitted for publication (Buscher-möhle, Verhey, Feudel, and Freund, 2007b). The main differences between the submitted manuscript and the present chapter are that in the present chapter Fig. 3.2 has been added and that Sect. 3.5.1 has been modified by adding a sentence. Additionally, several individual words have been rephrased.

identified an object, listeners can follow it even under adverse circumstances. Object formation may be guided by many different cues, one of which is amplitude modulation. It has been shown that the detectability of signals can depend on the correlation structure of modulations in an auditory stimulus (e.g. Nelken et al., 1999; Singh and Theunissen, 2003). As amplitude modulations due to sound production and sound propagation occur frequently in nature, this cue might be especially important for animals and humans (e.g. Langemann and Klump, 2001; Hofer and Klump, 2003; Verhey et al., 2003).

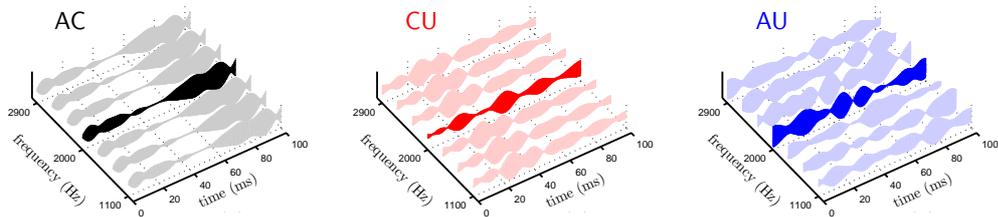


Figure 3.1: Schematic illustration of typical correlation conditions for the noise bands used in comodulation experiments. The flanking bands (FB) are light-colored while the signal band (SB) or on-frequency masker (OFM) are dark-colored. Left: FBs and SB/OFM have correlated amplitude modulations (all correlated, AC). Middle: FBs are comodulated, SB/OFM has uncorrelated amplitude fluctuations (co-uncorrelated, CU). Right: amplitude fluctuations of FBs and SB/OFM are uncorrelated (all uncorrelated, AU). Compare also Figs. 1.1 and 2.1.

Several experiments in auditory object formation are concerned with correlated amplitude modulations and their effect on signal detection. Two such experimental paradigms are comodulation detection difference (CDD) and comodulation masking release (CMR, see Table 3.1 for a summary of all abbreviations used in this chapter). In the first of these paradigms (e.g. Cohen and Schubert, 1987; Fantini and Moore, 1994; Wright, 1990; Hall et al., 2006), the signal to be detected is a narrow noise band (signal band) in the frequency domain masked by one or several additional noise bands (flanking bands). If the envelope of the signal band fluctuates in the same way as that of the flanking bands, detection thresholds are found to be higher than if the flanking bands share a common envelope while the signal band has a differing envelope. The difference in these thresholds is called the comodulation detection difference (CDD). For the case of more than one flanking band, three main correlation conditions may be distinguished: *(i)* all correlated (AC), if all envelopes are the same, *(ii)* all uncorrelated (AU) for mutually different envelopes, and *(iii)* co-uncorrelated (CU) if the flanking band envelopes are the same while the signal band has a different envelope (see Fig. 3.1). Generally, signal detection thresholds in CDD experiments are lowest in the

Table 3.1: Abbreviations used in this chapter.

Abbreviation	Meaning
AC	all correlated
AU	all uncorrelated
CDD	comodulation detection difference
CMR	comodulation masking release
CU	co-uncorrelated
ERB	equivalent rectangular bandwidth
FB	flanking band
OFM	on-frequency masker
SB	signal band
SPL	sound pressure level

CU condition. With this nomenclature, CDD is often defined by subtracting the AC threshold from the CU threshold, and therefore it is usually negative (McFadden, 1987).

Stimuli commonly used in CMR experiments consist of a narrow noise band (the on-frequency masker, OFM) masking a pure tone signal and one or several flanking noise bands serving as additional maskers. If the flanking bands are modulated in the same way as the OFM (which will be referred to as the AC condition according to the above terminology), then signal detection thresholds are lower than if flanking bands and OFM have differing amplitude modulations (in the following referred to as the AU condition). This threshold difference has been termed comodulation masking release (CMR, Hall et al., 1984). A second kind of CMR experiments exists in which there is only one on-frequency masker centered on the signal sinusoid. Common amplitude fluctuations in this case may be obtained by modulating the whole on-frequency masker with a lowpass noise. This condition can also be called the AC condition, while the unmodulated case can be termed the AU condition. With increasing OFM bandwidth, threshold differences between AC and AU condition tend to increase in these experiments, even if the masker bandwidth exceeds the width of a typical auditory filter. This finding has led to the hypothesis that across-channel processes may underlie the CMR effect (Hall et al., 1984), however, see Verhey et al. (1999). In the present work, CMR is defined as threshold difference between the AU and AC conditions (AU-AC, usually positive).

Several mechanisms have been hypothesized to account for CMR (see Verhey et al., 2003, for a review) and CDD (see Moore and Borrill, 2002, and references therein for further information). Generally, the explanations can be divided into within-channel and across-channel accounts of the ob-

served effects. This refers to the number of auditory channels involved in the processing of the stimuli (The inner ear can be viewed as a filterbank that analyzes sounds according to a row of auditory filters or channels).

The across-channel explanation for the psychophysical CMR experiments is that the auditory system compares the output of a channel that is centered on the signal to those centered on the flanking bands. Model realizations of such across-channel comparisons may be *(i)* correlation models, in which the outputs of several channels are cross-correlated with one another, *(ii)* equalization-cancellation models, in which the outputs of several channels are first equalized in overall level and then subtracted from each other, or *(iii)* dip listening models, in which the auditory system is assumed to be able to detect times of low amplitudes of the flanking band channels in order to improve the signal-to-noise ratio.

Dip listening, however, does not necessarily have to rely on across-channel processing. Mechanical suppression on the level of the cochlea and other mechanisms have been proposed to allow for dip listening also within one auditory channel (see Moore and Borrill, 2002; Ernst and Verhey, 2006). Another possibility of a within-channel explanation is that changes in the temporal waveforms may be registered by the auditory system and lead to different signal detection thresholds in the various correlation conditions. A quantitative within-channel model for CMR is proposed in Verhey et al. (1999). In this model, the most important stage is a spectral decomposition of the envelope within the auditory filter by means of a modulation filterbank located after the inner ear's frequency filterbank on the auditory pathway (Dau et al., 1997), the output of which is compared to a stored "image" for a suprathreshold signal by using cross correlation. This within-channel model can explain the experimental data in the accompanying experiments by an effective reduction of the modulation depth.

Possible explanations for psychophysical CDD findings are very similar to those used for CMR. Different qualitative across-channel and within-channel mechanisms have been proposed. Borrill and Moore (2002) and Moore and Borrill (2002) conclude that CDD is most likely mainly a within-channel effect. They introduce a quantitative within-channel model in which the auditory system can detect times during which the signal-to-masker ratio in the output of one channel is temporarily above a certain threshold. These times are combined to the cumulative detection time over the total duration of the stimulus. If this cumulative time is sufficiently large, then the signal is detected.

These effective models are quite elaborate and use technical approaches to explain psychoacoustical data. Their neuronal realization is, however, still not fully understood. Wide-band inhibition in the cochlear nucleus has been

proposed as the neural mechanism underlying CMR, where the response of neurons that are excited by the signal frequency of the stimulus is inhibited by neurons that react to a wider range of frequencies around the signal frequency (Pressnitzer et al., 2001; Meddis et al., 2002; Neuert et al., 2004). This mechanism leads to a stronger firing rate response to the signal during masker dips in AC conditions than in AU or CU conditions. Another proposed neuronal explanation for CMR is suppression of firing rate locking to the masker envelope (Nelken et al., 1999; Las et al., 2005), which means that the sinusoidal signal prevents the stimulated neurons from locking to the masker envelope and thus allows for a better signal detectability in comodulated noise (corresponding to the AC condition) than in unmodulated noise (corresponding to the AU condition). The neuronal networks necessary for these explanations of CMR require at least a few excitatory and inhibitory synaptic connections to enable the proposed responses.

For CDD, a much simpler neuronal model based on neurophysiological recordings was proposed by Buschermöhle et al. (2006), which is explained in Chapter 2. The model involves only an excitatory stage and uses the mean value of the compressed envelope of a stimulus filtered by a single auditory channel as a detection cue. Although all the above mentioned neuronal models can account for general aspects of the respective psychophysical experiments, their predictions have not yet been compared directly to psychophysics.

To our knowledge, none of the models for CMR or CDD has aimed at explaining both effects at the same time, and the psychophysical models so far have not been directly linked to the physiology. The aim of the present work is to show that a very simple neuronal model can be applied to various psychophysical CDD *and* CMR experiments. The physiology-based within-channel model introduced by Buschermöhle et al. (2006) will be extended in order to account for both effects. For this purpose, first the basic steps of the model will be described, then its main mathematical expressions will be introduced, and subsequently its applicability to CDD and CMR experiments will be demonstrated. Finally, there will be a comparison of the model's predictions with data from the literature.

In the model, it is assumed that the mean firing rate of neuronal populations covaries with the mean compressed envelope of the filtered stimulus. Such a locking of neuronal firing rates to the envelopes of auditory stimuli is described e.g. in Schreiner and Urbas (1988) and Joris et al. (2004). In Chapter 2, an expression for the time and trial expectation value of the compressed envelope of the filtered stimulus is derived for those CDD stimuli that were used for the corresponding physiological experiments. With this expression, the characteristics of neuronal firing rates could be simulated (see

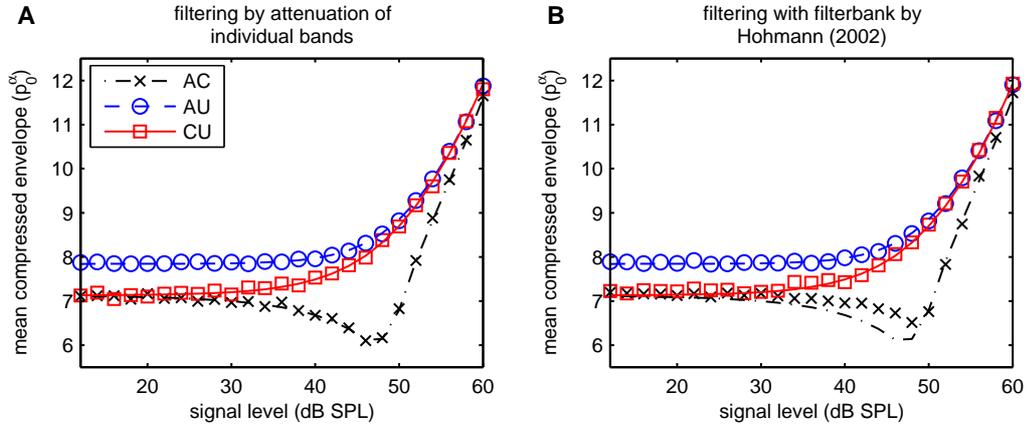


Figure 3.2: Comparison of mean compressed envelope values predicted by the model (lines) and obtained from simulations (symbols). The model predictions in both subplots are the same, only the simulated data points marked by the symbols differ. **A:** the stimulus consisting of masker bands and signal band was filtered by attenuating the individual bands, keeping the phase information intact. **B:** The stimulus was filtered with a filter from the gammatone filterbank described in Hohmann (2002), leading to phase distortions and therefore to deviations between model predictions and numerically obtained data. (Parameters: $\alpha = 0.35$, signal frequency $f_0 = 2$ kHz, $\gamma = 2.5$; stimulus setup as described in Chapter 2, overall level of each of the six masker bands: 50 dB SPL.)

Bee et al., 2007 and Chapter 2 for details on the experimental setup). Here, the calculations are generalized to psychophysics and to describe CMR as well.

3.3 MODEL STRUCTURE

The model is organized in several steps. First, the incoming stimulus is filtered by applying a bandpass filter that is centered on the signal frequency. In order to allow for analytical calculations, each noise band is attenuated as a whole according to the response of a filter centered on the signal using the magnitude transfer function of the gammatone filterbank described in Hohmann (2002).

This procedure means that a phase preserving filter is assumed. For the model results presented here, the difference between this form of frequency-dependent attenuation and filtering may be neglected: numerical results with the original filters do not differ qualitatively from the analytical approximations (see Fig. 3.2). Quantitative differences may be compensated by choosing adequate parameters.

After frequency filtering, the trial averaged envelope of the resulting stim-

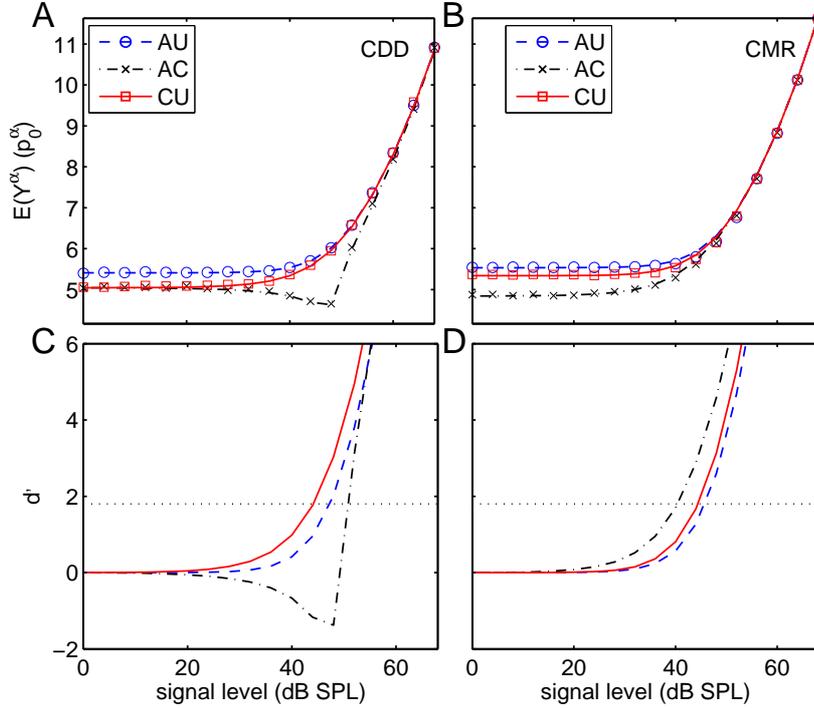


Figure 3.3: Mean compressed envelope values for CDD (A) and CMR (B) stimuli as well as the corresponding index of discriminability (d') curves (C and D). Markers indicate simulated data, while lines are derived from Eqns. 3.7, 3.8 and 3.9. The overall level is set to 50 dB SPL for each flanking band, and the overall level of the on-frequency masker is 40 dB SPL. Compression $\alpha = 0.3$. Center frequency $f_0 = 2.0$ kHz. Flanking bands (and on-frequency masker in case of CMR) have a bandwidth of 100 Hz and are centered at 1.7 kHz, 1.85 kHz, 2.0 kHz, 2.15 kHz, and 2.3 kHz. Each band is attenuated as a whole according to the magnitude transfer function of a gammatone filter centered at 2.0 kHz (bandwidth parameter $\gamma = 1$). The dotted horizontal line in panels C and D indicates the decision criterion $D = 1.8$. The signal levels at which the d' -curves cross this criterion are the signal detection thresholds. ($\sigma = 0.3$)

ulus is calculated. Then a compressive nonlinearity is applied to the envelope by raising the envelope to the power of α with $0 < \alpha < 1$. To finally get an expression for the mean of this compressed envelope across time and trials, the statistics of the analytical signal (Gabor, 1946) is analyzed.

The stimuli in CDD as well as in flanking band CMR experiments consist of a number of noise bands centered at certain frequencies and in the case of CMR an additional single sinusoid. Numbering the flanking noise bands from 1 to K and giving the on-frequency masker (or accordingly the signal

band) the number 0, the analytical signal can be written as:

$$s(t) = d_0 b e^{2\pi i f_0 t} + \sum_{k=0}^K a_k d_k \sum_{n=-N}^N e^{i[2\pi(f_k + n\Delta\nu)t + \phi_{k,n}]}, \quad (3.1)$$

where each noise band is regarded as a sum of $2N + 1$ pure tones with random phases. The factors d_k produce the attenuation due to the filtering process. The envelope of the filtered stimulus is $|s(t)|$, while the filtered stimulus itself is $\text{Re}(s(t))$. Each individual noise band is centered at f_k and has a bandwidth of $2N\Delta\nu$. The noise bands are composed of individual sinusoids with amplitudes a_k , while the sinusoidal signal at f_0 for CMR-experiments has the amplitude b and is attenuated by the filter factor d_0 (for CDD-experiments, $b = 0$). The phases $\phi_{k,n}$ distinguish the different correlation conditions: In the AU condition, the $\phi_{k,n}$ are independent and uniformly distributed in $[0; 2\pi]$. In the AC condition, all phases within one band (constant k , varying n) are random, while the same set of phases is used for all the different bands. And finally, in the CU condition, the phases of the flanking bands (i.e. $k \in \{1, \dots, K\}$) are random within one band but the same for different bands, while the phases of the signal band (i.e. $k = 0$) are independent of the phases of the FBs and randomly distributed.

The trial average of the squared envelope across phases (i.e. the ensemble average) can be calculated and simplified by splitting the absolute square $|s(t)|^2$ into a sum of $\text{Re}^2(s(t))$ and $\text{Im}^2(s(t))$ and by using addition theorems. Straightforward calculations lead to:

$$\begin{aligned} \langle |s(t)|^2 \rangle_\phi &= d_0^2 b^2 + (2N + 1) \sum_{k,k'=0}^K a_k d_k a_{k'} d_{k'} \cdot \\ &\quad \langle \delta(\phi_{k,0}, \phi_{k',0}) \rangle_\phi \cos(2\pi|f_k - f_{k'}|t). \end{aligned} \quad (3.2)$$

This expression can be split into the three correlation conditions. The simplest case is the AU condition, where one can write $\langle \delta(\phi_{k,0}, \phi_{k',0}) \rangle_\phi = \delta(k, k')$:

$$\langle |s(t)|^2_{\text{AU}} \rangle_\phi = d_0^2 b^2 + (2N + 1) \sum_{k=0}^K a_k^2 d_k^2. \quad (3.3)$$

In the AC condition, the δ -term is always one, such that

$$\begin{aligned} \langle |s(t)|^2_{\text{AC}} \rangle_\phi &= d_0^2 b^2 + (2N + 1) \sum_{k,k'=0}^K a_k a_{k'} d_k d_{k'} \cdot \\ &\quad \cos(2\pi|f_k - f_{k'}|t). \end{aligned} \quad (3.4)$$

Finally, in the CU condition, the δ -term does not vanish for all $k, k' \neq 0$ and if $k = k' = 0$. Thus, one gets

$$\begin{aligned} \langle |s(t)|_{\text{CU}}^2 \rangle_\phi &= d_0^2 b^2 + (2N + 1) d_0^2 a_0^2 + \\ &\quad (2N + 1) \sum_{k, k'=1}^K a_k a_{k'} d_k d_{k'} \cdot \\ &\quad \cos(2\pi |f_k - f_{k'}| t). \end{aligned} \quad (3.5)$$

In Appendix 3.7, Eqns. 3.3–3.5 are rewritten in a notation using the overall levels of the noise bands. The differences between the three correlation conditions can be seen in the interference terms $\cos(2\pi |f_k - f_{k'}| t)$.

Averaging the above equations over time will yield the same result in all three correlation conditions. The important step for getting quantitative differences in the temporal averages is to not consider the mean squared envelope but the mean compressed envelope (see also van de Par and Kohlrausch, 1998b; Verhey et al., 2007), which will be done in the following.

In order to proceed, we first consider the case of $b = 0$ (i.e. CDD). Here, the stimulus consists of noise bands only. This means that for large N , the distribution of squared envelope values may be approximated by an exponential distribution (Lawson and Uhlenbeck, 1950). The mean of this distribution at time t is given by $\mu^2(t) := \langle |s(t)|^2 \rangle_\phi$. (We define the term $\mu^2(t)$ for $b = 0$.) Now let Y denote the random variable describing the value of the stimulus envelope. Then for any $\alpha > 0$ the time dependent expectation value $E_t(Y^\alpha)$ can be computed from the exponential distribution of Y^α as

$$E_t(Y^\alpha) = \Gamma\left(\frac{\alpha + 2}{2}\right) |\mu(t)|^\alpha \quad (3.6)$$

Here, $\Gamma(\cdot)$ denotes the complete gamma function (e.g. Weisstein, 2002). In a last step, the time- and ensemble averaged value of the compressed envelope is given by integrating $E_t(Y^\alpha)$ over time. If the f_k share a common integer divisor \tilde{f} , the integration only needs to be done over one period $T = 1/\tilde{f}$. Else, one needs to take the limit of $T \rightarrow \infty$ for getting the expectation value:

$$E(Y_{\text{CDD}}^\alpha) = \Gamma\left(\frac{\alpha + 2}{2}\right) \frac{1}{T} \int_0^T |\mu(t)|^\alpha dt. \quad (3.7)$$

Eqns. 3.6 and 3.7 are derived for the CDD case ($b = 0$). In the case of CMR the equations will still be a reasonable approximation for small b (i.e. $b \ll a_0$). But with increasing amplitude of the sinusoidal signal in the CMR stimuli, the distribution of the squared envelope values will not be exponential anymore. For $b \gg a_0$ the sinusoid will dominate the envelope

Table 3.2: Model parameters used in this chapter.

Parameter	Range	Description
α	0.1–0.6	compressive exponent (no units)
σ	0.1–0.35	internal error (in p_0^α , compressed reference pressure)
γ	0.8–5.0	filter width (in ERB)

which means that the distribution of Y^2 can be approximated by a δ -peak at $d_0^2 b^2$, such that $E(Y^\alpha) \approx d_0^\alpha b^\alpha$. As the calculations leading to Eq. 3.7 cannot be carried out without making an assumption about the distribution of squared envelope values, we choose to approximate $E(Y^\alpha)$ by getting the right asymptotic behavior for small b and for large b within one expression. The simplest way to approximate this asymptotic behavior is to use the prefactor $\Gamma((\alpha + 2)/2)$ only for $\mu^2(t)$, but not as a prefactor for the term describing the signal sinusoid (i.e. the term including b):

$$E(Y_{\text{CMR}}^\alpha) = \frac{1}{T} \int_0^T \left(d_0^2 b^2 + \Gamma\left(\frac{\alpha+2}{2}\right)^{\frac{2}{\alpha}} |\mu^2(t)| \right)^{\frac{\alpha}{2}} dt. \quad (3.8)$$

This expression has the desired asymptotic behavior.

A comparison of the predictions made by Eqns. 3.7 and 3.8 with realizations of stimuli from simulations is shown in Fig. 3.3 A and B. Although for simplicity constant amplitudes of the sinusoids making up the stimuli have been assumed, the calculations hold with good accuracy also for stimuli with Rayleigh-distributed amplitudes, which are frequently used in experimental setups. They are also reasonably accurate for describing stimuli where the noise bands are generated by multiplying lowpass noise with sinusoids centered at the noise band center frequencies. This means that for explaining the general effects the model calculations can also be applied to CDD and CMR stimuli from different authors that have been generated in slightly different ways. A remarkable feature of the model curves shown in Fig. 3.3 A is that their general shape is very similar to the characteristics of neuronal firing rates measured in the avian auditory forebrain during presentation of CDD stimuli (Buschermöhle et al., 2006; Bee et al., 2007). Differences of the curves for different correlation conditions result from interference between the components for correlated phases (see the cosine terms in Eqns. 3.4 and 3.5).

3.3.1 Mechanism of signal detection

A basic signal detection scheme can be envisaged by assuming that the time averaged compressed envelope of a filtered sample stimulus is represented somewhere in the auditory system, possibly by the firing rate of a population of neurons. This estimate y is a random variable and will have an error σ associated with it which is due to variability within the stimulus as well as variability in its neural representation. In the following, this error σ is assumed to be constant (in particular, independent of signal and masker level), which allows for using the d' -measure from signal detection theory (Green and Swets, 1966):

$$d'(L_S) = \frac{y(L_S) - y(-\infty)}{\sigma}, \quad (3.9)$$

where $y(L_S)$ denotes the estimate of the mean compressed envelope value of the filtered stimulus when the signal level is at L_S and $y(-\infty)$ denotes the same estimate when the signal is absent. The larger the value of d' , the further apart are the two distributions of y with and without signal. If d' exceeds a predefined criterion $D = 1.8$, then the signal is said to have been detected, while for $d' < D$, the signal is not detected. The signal detection threshold can be defined as the signal level at which d' reaches D . Note that the parameters σ and D together only comprise one free parameter: a change in the decision criterion D can be equivalently expressed as a change in the internal error σ , which essentially rescales the d' axis.

With this signal detection scheme, one can identify the reason for CDD and CMR in Fig. 3.3. For CDD, consider the AC and CU curves in Fig. 3.3 A. Calculating d' means shifting them to zero and rescaling them by $\frac{1}{\sigma}$ (see Fig. 3.3 C). Because the AC curve is always beneath the CU curve, the former will cross the detection criterion $d' = D$ at a higher signal level than the latter, which directly corresponds to the CDD effect. For CMR (Fig. 3.3 B), the AU and AC curves are both monotonically increasing and have the same asymptotic behavior for $L_S \rightarrow \infty$, but values for vanishing signal are not the same. Calculating d' corresponds to shifting both curves vertically relative to each other, and therefore, the AC curve will always cross the detection criterion at lower signal levels than the AU curve (see Fig. 3.3 D), i.e. a CMR is predicted.

3.3.2 Model parameters

Up to now, two free model parameters have been introduced: the compression α and the internal error σ . The third and final model parameter γ

describes the width of the auditory gammatone filter in ERB. The ERB (equivalent rectangular bandwidth) of a bandpass filter with arbitrary amplitude response is the bandwidth of a corresponding filter with rectangular amplitude response that has the same peak response and passes the same total amount of power. The relation between center frequency f_c and ERB bandwidth for the human auditory system is given by the following empirical formula from Glasberg and Moore (1990) for $\gamma = 1$:

$$ERB(f_c) = \gamma \cdot (24.7 \text{ Hz} + 0.1079f_c). \quad (3.10)$$

The prefactor γ is introduced here to be able to change the filter width. For exploring the model's ability to explain different CDD and CMR experiments, these three parameters will be adjusted within reasonable ranges in the following. The parameter meanings and their ranges are summarized in Table 3.2. The influence of the different parameters is discussed in Sect. 3.5. The compression in the human auditory system can vary roughly between $\alpha = 0.1$ and $\alpha = 0.8$ (for a review, see Bacon et al., 2003). The internal error σ is a quantity that is hard to measure, and therefore, its reasonable range for the human auditory system is not known. We regard σ as an adjustable parameter which should be positive and smaller than the dynamic range of mean compressed envelopes that arises at sound pressure levels between 0 dB and 100 dB. The filter bandwidth parameter γ should be close to one for normal hearing humans. Values up to about 1.5 may be realistic for within-channel processing of stimuli (see Glasberg and Moore, 1990). If γ is larger than that, then the processing must be regarded as across-channel processing.

3.4 RESULTS

The very simple model introduced here has only three free parameters and has been derived from a model for neuronal firing rates obtained in CDD experiments. The basic modeling steps frequency filtering, envelope extraction, and compression are directly motivated by peripheral auditory processing. Only temporal averaging and decision making are presumably central processes. Despite its simplicity, the model will now be applied to different CDD and CMR experiments described in the literature. Model and experimental results are compared in order to determine the scope of the proposed model. All relevant parameters for the experiments modeled in the present study are summarized in Table 3.3.

Table 3.3: An overview of the experimental setups and respective modeling parameters discussed in this chapter. The experimentally varied quantities are printed in bold face.

Reference	Paradigm	Signal frequency	FB frequencies	Bandwidth	Noise band level	α	σ	γ	Remarks
Borrill and Moore (2002)	CDD	1.5 kHz	$\Delta f = \pm 0.1$... ± 1.4 kHz	20 Hz	78 dB SPL	0.3	0.1	1	-
Cohen and Schubert (1987)	CDD	0.2 ... 6.0 kHz	1.0 kHz	100 Hz	73 dB SPL	0.6	0.1	1	-
McFadden (1987)	CDD	2.5 kHz	1.5 ... 3.5 kHz	100 Hz	70 dB SPL	0.6	0.1	1	-
Schooneveldt and Moore (1987)	CMR	2.0 kHz	1.0 ... 3.0 kHz	25 Hz	67 dB SPL	0.2	0.1	5	-
McFadden (1987)	CMR	2.5 kHz	1.5 ... 3.5 kHz	100 Hz	70 dB SPL	0.2	0.1	5	-
Hall et al. (1984)	CMR	1.0 kHz	0.7 ... 1.3 kHz	100 Hz	60 dB SPL	0.2	0.1	5	-
McFadden (1987)	CDD	2.5 kHz	2 or 4 bands at 1.5 ... 3.5 kHz	100 Hz	70 dB SPL	0.2	0.1	1	-
McFadden (1987)	CMR	2.5 kHz	2 or 4 bands at 1.5 ... 3.5 kHz	100 Hz	70 dB SPL	0.2	0.1	1.4	-
Borrill and Moore (2002)	CDD	1.5 kHz	0.9 kHz, 2.1 kHz	20 Hz	53 ... 83 dB SPL	0.3	0.2	0.8	-
Ernst and Verhey (2005)	CMR	8.0 kHz	1.0 kHz	100 Hz	OFM level: 20 ... 60 dB SPL	0.3	0.1	1	off-frequency listening
Hall et al. (1984)	CMR	1.0 kHz	1.0 kHz	100 ... 700 Hz	40 dB SPL spectrum level	0.25	0.35	1.2	-

3.4.1 Frequency spacing between signal and flanking bands for CDD

Several experimental studies discuss the effect of frequency distance between flanking bands and the signal band for CDD (e.g. Borrill and Moore, 2002; Cohen and Schubert, 1987; McFadden, 1987). In the present study, the model is applied to stimulus setups comparable to the ones described in these three studies. The leftmost column in Fig. 3.4 compares model predictions with experimental data from Borrill and Moore (2002), where the signal band was flanked symmetrically by two noise bands, at a distance Δf above and below the signal frequency, respectively. In the central column of Fig. 3.4, model and experiment are compared for data published in Cohen and Schubert (1987). There, the stimulus consisted of a flanking band at a constant center frequency and signal band at varying distances Δf from the flanking band. In the third column of Fig. 3.4, signal detection thresholds for an experiment described in McFadden (1987) are plotted. Here, the stimulus consisted of one signal band at a constant center frequency and a flanking band with varying center frequency.

The general shape of the model thresholds for the AC and CU conditions depending on Δf in all three experiments qualitatively reproduces the shape of the experimental thresholds. The difference between AC and CU thresholds (the CDD) decreases for increasing frequency spacing but does not reach zero for the largest Δf . The thresholds are determined by the form of the auditory filter since the filter used for signal detection is always the one centered on the signal band. The signal band will be detected as soon as its influence on the total filter output dominates the mean compressed envelope, which happens at signal levels close to the level of the attenuated flanking band. The clearest deviations between model and experiment can be found for Δf close to 0 Hz. There, the model predicts quite large CDDs while in experiments the CDDs are relatively small². The plot for the experiments by Cohen and Schubert (1987) is not symmetrical to $\Delta f = 0$ Hz. This reflects the fact that for each signal frequency a different filter is picked, and that filters with increasing signal frequencies have larger bandwidths (cf. Eq. 3.10). In contrast, the plot for data from McFadden (1987) is largely symmetrical with respect to $\Delta f = 0$ Hz because the signal frequency and therefore the auditory filter does not change. Here, the thresholds reflect the shape of the symmetrical filter.

All three experiments could be modeled reasonably well with very similar

²Note that for $\Delta f = 0$ Hz, the AC condition needs to be treated separately (see Appendix 3.7), because here signal band and masker band are exactly the same except for level differences. We do not explicitly treat this case here because it is not important for the general model results.

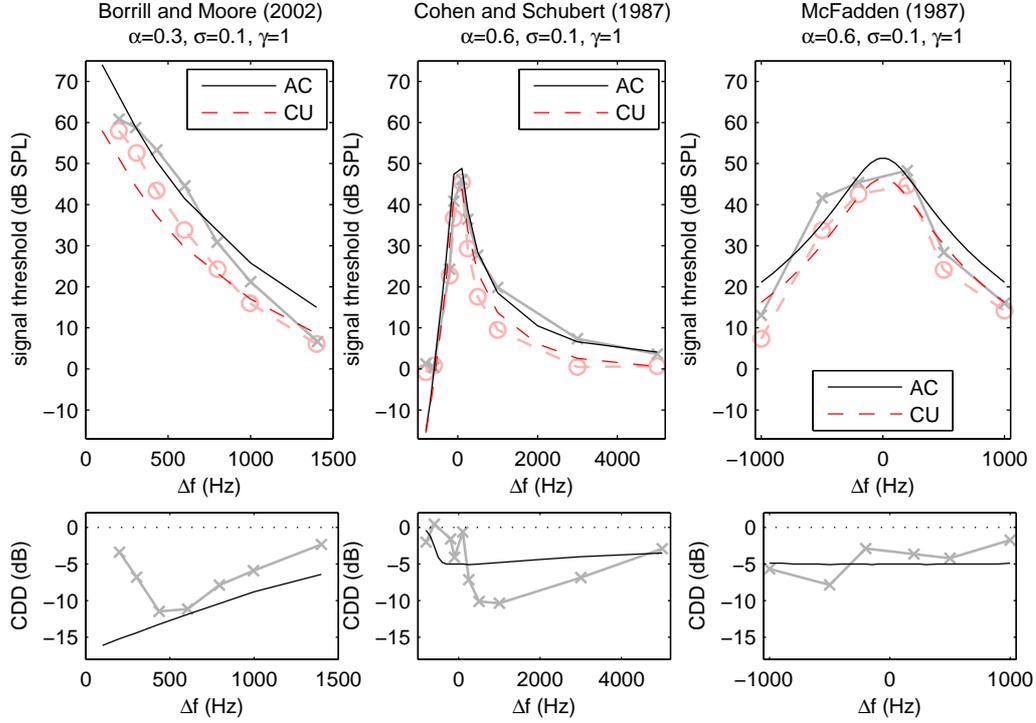


Figure 3.4: Effect of frequency spacing between signal and flanking band on CDD. Dark thin lines mark model results. Light broad lines with symbols indicate experimental data. The three columns show data from three different CDD experiments. References and model parameters used are indicated in the figure titles. The top row shows signal detection thresholds, the bottom row shows the corresponding threshold differences (CDDs). Experimental parameters are summarized in Table 3.3.

parameters. It is noteworthy that the filter bandwidth is at a value that is typical for the human ear, indicating that the whole CDD effect may be due to peripheral within-channel processes.

3.4.2 Frequency spacing between the signal and one flanking band for CMR

The effect of frequency distance between one flanking band and the on-frequency masker for CMR is discussed amongst others in Schooneveldt and Moore (1987), McFadden (1987), and Hall et al. (1984). The model is compared to the results reported in these studies in Fig. 3.5. In that figure, the left column shows data from Schooneveldt and Moore (1987), where the signal frequency was kept constant and one flanking band was varied in its center frequency. The frequency distance between the flanking band's center frequency and the signal frequency is denoted as Δf . The second column in

Fig. 3.5 shows data from McFadden (1987), where the on-frequency masker and the signal had a constant center frequency while the flanking band was centered at varying distances Δf from the signal. In the rightmost column of Fig. 3.5, experimental results from Hall et al. (1984) are compared to model predictions. The stimuli here consisted of an on-frequency masker and a signal at a constant frequency and one flanking band at a frequency distance Δf . The case $\Delta f = 0$ Hz was treated as a special case in that experiment: For that case there was only the on-frequency masker present (without the flanking band) which means that there was no difference between the AC and AU correlation conditions and therefore both signal detection thresholds are the same in the model as well as in the experiments.

A general finding for all three experiments is that thresholds in the AU case increase for decreasing Δf , while thresholds in the AC case behave in the opposite way. For large frequency separations, the threshold differences (i.e. the CMR) decrease, which can also be seen in the experimental data. The general shape of the depicted model curves can be understood by considering that for large frequency separations, the flanking band is completely filtered out by the auditory filter and therefore only the on-frequency masker determines the signal threshold. The closer the flanking band is to the on-frequency masker, the more it influences the on-frequency masker. In the AU case, it increases the power that falls into the auditory filter and therefore rises thresholds, while in the AC case, the flanking band causes beating with the on-frequency masker and therefore leads to reduced mean compressed envelopes and facilitates signal detection.

The parameters used for modeling the three experiments are exactly the same, which means that in spite of slightly varying experimental conditions, the model can explain these CMR experiments without the need for fine-tuning. An important note needs to be made concerning the parameter γ determining the bandwidth of the auditory filter. The value $\gamma = 5$ is well above the realistic range for typical auditory filters. This means that CMR is presumably not a within-channel process as it was the case for CDD in Sect. 3.4.1. However, the model described here does not rely on classical cross-channel processes as they were explained in Sect. 3.2. This point will be discussed further in Sect. 3.5.1.

3.4.3 Number of flanking bands for CDD and CMR

It has been pointed out that the amount of CDD as well as CMR depends on the number of flanking bands (McFadden, 1987, Wright, 1990 for CDD, and Hall et al., 1984, McFadden, 1987 for CMR). This effect is investigated for

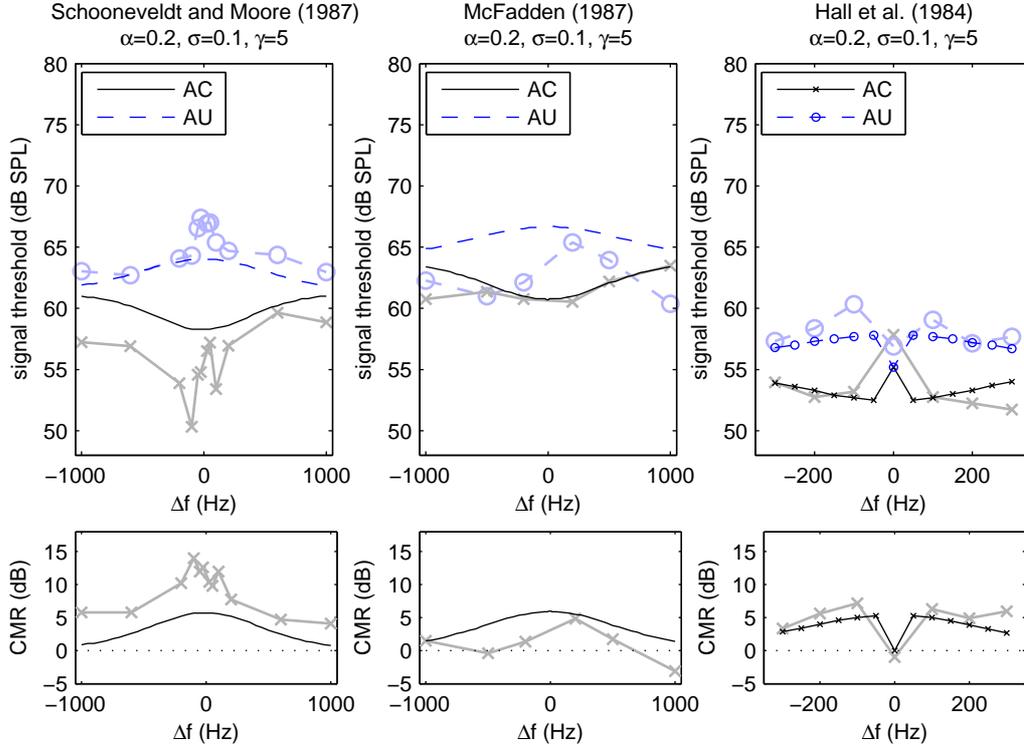


Figure 3.5: Effect of frequency spacing on CMR. Dark thin lines mark model results. Light broad lines with symbols indicate experimental data. The three columns show data from three different CMR experiments. References and model parameters used are indicated in the figure titles. The top row shows signal detection thresholds, the bottom row shows the corresponding threshold differences (CMRs). In all experiments, the signal was a pure sine tone at a constant signal frequency, and there were two noise bands present: the on-frequency masker centered on the signal frequency and the flanking band with varying frequency distance from the signal Δf . The signal frequency, noise bandwidth and noise band levels were different for the different experiments (see Table 3.3). Model predictions for the experiment by Hall et al. (1984) were obtained only for certain frequency distances indicated by the symbols.

the model by setting up the stimuli similar to those used in McFadden (1987), where the consequences of adding further flanking bands are investigated for both CDD and CMR. The model results are compared to experimental data in Fig. 3.6. In the experiments, only the AC and CU conditions were considered, which is why the model results are restricted to these two correlation conditions.

For CDD (Fig. 3.6 A and C), it is observed in the model and in the experiments that both thresholds are higher if the flanking bands are closer to the signal band. This is the case because by shifting the flanking bands closer

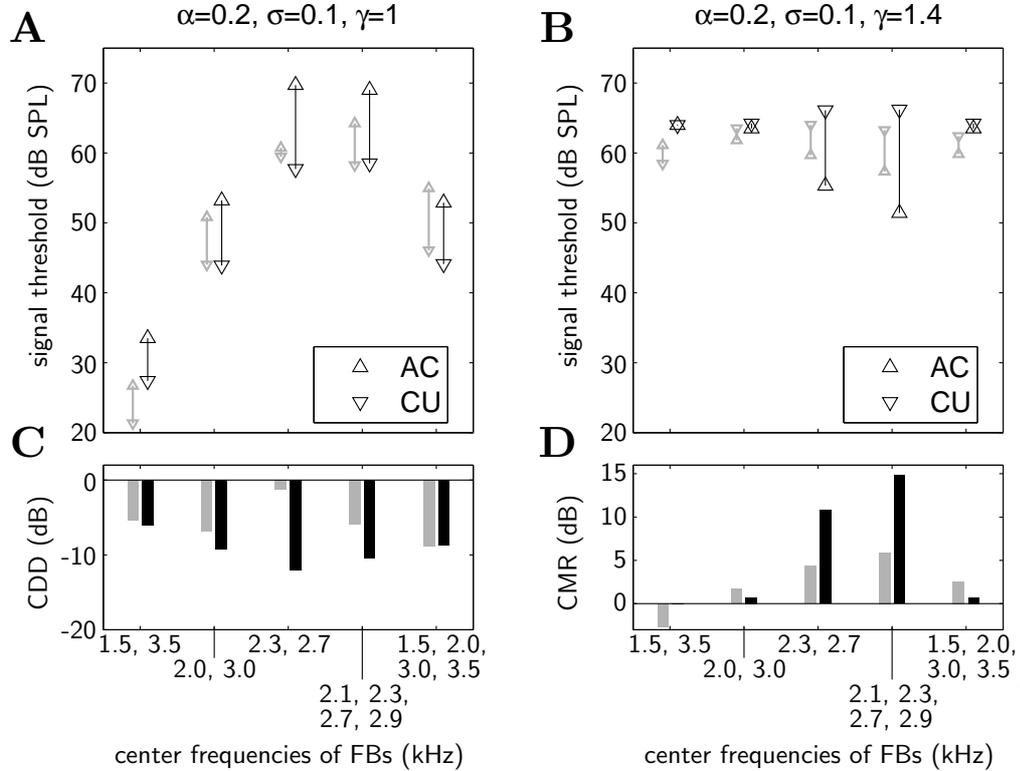


Figure 3.6: Effect of number of bands on CDD (**A**, **C**) and CMR (**B**, **D**). Dark triangles mark signal detection thresholds as determined from the model. Light triangles indicate experimental data from McFadden (1987). **C**: CDD, gray bars: experimental data, black bars: model. **D**: CMR. Model parameters are specified in the titles. Experimental parameters are summarized in Table 3.3.

to the signal band, they are less attenuated due to filtering and therefore contribute to increased thresholds. The amount of CDD in the model grows the closer the flanking bands come to the signal band. This is not reflected in the experimental data. Also the experimental finding that by adding two further noise bands, the amount of CDD increases is not reflected in the model. These discrepancies between model and experiment are consistent with those found in Sect. 3.4.1, where the CDDs for very small Δf increased in the model while they decreased in the experiments. Still, the general characteristics of the experimental thresholds (Fig. 3.6 A) are reasonably close to the model predictions.

Comparison of model predictions and experimental results for CMR is shown in Fig. 3.6 B and D. Here, the model predicts considerable CMR only for the cases where the flanking bands are closest to the on-frequency masker. This generally agrees with the experimental data. The model thresholds

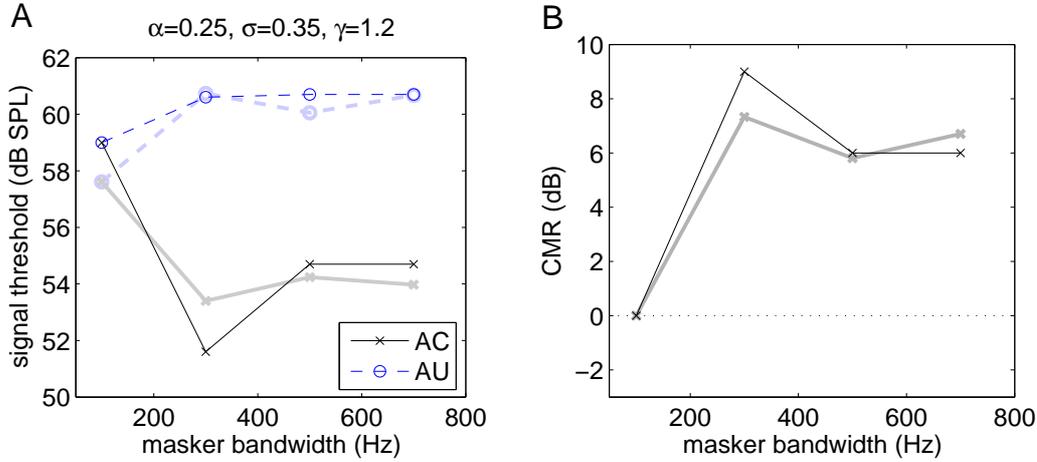


Figure 3.7: Effect of masker bandwidth on CMR. Dark lines mark model results. Light lines indicate experimental data from Hall et al. (1984). The on-frequency masker is composed of up to seven noise bands with a bandwidth of 100 Hz each, and the bandwidth is increased by symmetrically adding further noise bands to the OFM. **A:** Signal thresholds, **B:** Threshold differences.

also are in the same range as the experimental thresholds, but the measured threshold differences (CMRs) are quantitatively different from the model predictions, although the general tendency of increasing CMR by adding further noise bands is reflected by the model.

A further experiment, described in Hall et al. (1984), investigates the influence of noise bandwidth on CMR by symmetrically adding 100 Hz wide noise bands to the right and left of the on-frequency masker. The additional noise bands can either have the same envelope fluctuations as the first on-frequency masker (AC) or independent envelope fluctuations (AU). This experiment can therefore also be seen as an experiment concerning the number of noise bands. Comparison of the data from Hall et al. (1984) with model predictions is shown in Fig. 3.7 A and B.

There, the on-frequency masker bandwidth is increased from 100 Hz over 300 Hz and 500 Hz to 700 Hz by repeatedly adding two noise bands of a bandwidth of 100 Hz above and below the original on-frequency masker. The model thresholds show a behavior which resembles that of the experimental CMR thresholds closely. The CMR does not change strongly for on-frequency masker bandwidths larger than 500 Hz because further masker power is only added at frequencies outside the filter which does not influence the envelope statistics significantly.

The parameters used for modeling the CDD and CMR experiments in this section are similar. The most important difference is the filter bandwidth γ . Again, for CMR, larger filters were needed than for CDD. However, all

parameter values can still be viewed as modeling within-channel processes.

3.4.4 Influence of noise band level

A further experimental parameter which may affect signal detection thresholds is the level of the noise bands. For CDD, the dependence of signal thresholds on flanking band level was explored in Borrill and Moore (2002). In that study, the signal band was centered between two flanking bands. Experimental and model data for this setup are compared in Fig. 3.8 A and C. One finds that signal thresholds generally increase with rising flanking band level. In the experiments, AC thresholds are generally higher than CU thresholds. The model explains this order of thresholds as well as the amount of CDD correctly.

There are several studies involving the dependence of CMR on noise band level. The dependence of CMR on the level of one flanking band, keeping the on-frequency masker level constant, is investigated in Schooneveldt and Moore (1987). In Moore and Shailer (1991), the dependence of CMR on the level of several flanking bands is examined keeping the on-frequency masker level constant or varying it with the flanking bands. In Cohen (1991) and Ernst and Verhey (2005) the influence of on-frequency masker level on the detection of the signal sinusoid is analyzed while the overall level of one flanking band is kept constant. As the last two studies report a CMR over a range of several octaves, we choose to compare the model with the data of the latter study, Ernst and Verhey (2005).

As can be seen from Fig. 3.5, the basic model setup yields significant CMRs only for frequency distances of up to 1000 Hz between on-frequency masker and flanking band, which is considerably less than the 7 kHz separation between the two bands here. Therefore, the model needs to be changed from using the filter centered on the signal frequency (on-frequency listening) to using the filter which attenuates flanking band and on-frequency masker such that their levels *after filtering* are the same. This corresponds to an off-frequency listening strategy (Patterson and Nimmo-Smith, 1980; O'Loughlin and Moore, 1981) and means that a different filter is picked from the filterbank than the one centered on the signal. If the filter used in the model is chosen in such a way, then beating between on-frequency masker and flanking band after filtering has a big effect on the envelope statistics in the AC condition and thus facilitates signal detection in the AC condition. Now, the central frequency of the filter used for signal detection depends on the levels of on-frequency masker and flanking band. The modeling results for on- and off-frequency listening are plotted in Fig. 3.8 B and D. One finds a good cor-

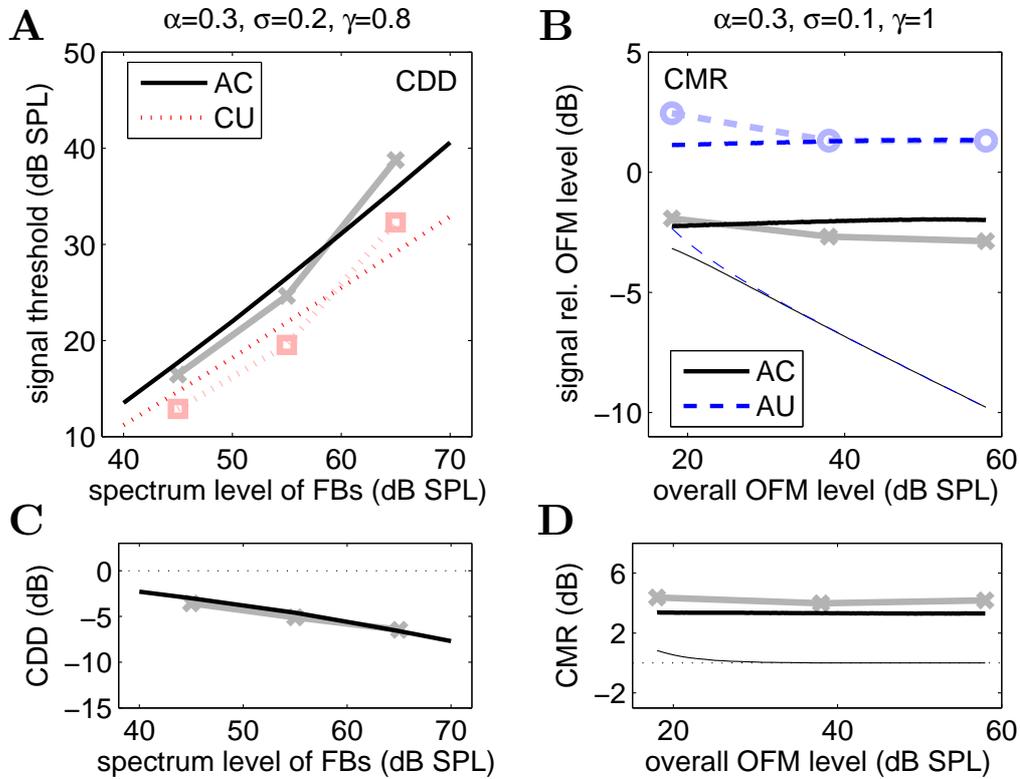


Figure 3.8: Effect of level of noise bands on CDD (A, C) and CMR (B, D). **A:** CDD; dark thick lines mark signal detection thresholds as determined from the model, light thick lines indicate mean experimental data from Borrill and Moore (2002). The signal band is centered between two flanking bands. **B:** CMR; light thick lines indicate data from Ernst and Verhey (2005). Dark thick lines represent model results for off-frequency listening. Thin lines show model results for on-frequency listening (filter centered on the signal frequency). The flanking band is located 7 kHz below the on-frequency masker at 8 kHz. **C, D:** Threshold differences.

responsiveness between the off-frequency listening model and the experimental data: The amount of CMR is nearly independent of on-frequency masker level.

Although the thresholds determined by the on-frequency listening model are lower than those predicted by the off-frequency listening model and therefore represent the better signal detection strategy, the performance of the auditory system in this experiment is better described by the off-frequency listening model. This is not the case for the previously modeled experiments.

3.5 DISCUSSION

In this chapter it is proposed that main features of CDD and CMR experiments can be understood by registering changes in the mean compressed envelope of a stimulus filtered by a single auditory channel. The model consists of the five stages frequency filtering, envelope extraction, compression, averaging, and signal detection, which may be realized at very early stages of the auditory pathway. There are three free parameters for the model: the filter bandwidth γ , the compressive exponent α , and the internal error σ . When applying the model to experiments from the literature, these model parameters are adjusted within reasonable ranges to demonstrate the scope of the model.

3.5.1 Filter bandwidth γ

The bandwidth of the auditory filters is usually kept close to the values given in Glasberg and Moore (1990) (corresponding to $\gamma \approx 1$ in Eq. 3.10). This choice is valid for most normal hearing subjects. The good correspondence between data and model predictions indicates that CDD and CMR effects in most of the experiments are due to within-channel effects. Only for the CMR experiments investigating the effect of frequency spacing between on-frequency and flanking bands, a standard auditory filter bandwidth cannot account for the experimental thresholds. By increasing γ , the present within-channel model can also explain across-channel contributions to CMR. This finding is comparable to suggestions by Berg (1996), who proposed that some CMR experiments can be explained by an analysis of the output of only one broadly tuned filter. The general discussion of comodulation experiments in the literature indicates that CDD experiments can be accounted for by within-channel processes (see Borrill and Moore, 2002), while CMR also needs across-channel processing (e.g. Verhey et al., 2003). This is in agreement with the findings in the present study. Compared to other models, the presented model has the advantage that it does not involve suppression (Ernst and Verhey, 2006), inhibition (Pressnitzer et al., 2001), or analysis of temporal features (Verhey et al., 1999; Nelken et al., 1999) but only sequential excitatory feed-forward processing to account for within- as well as across channel processes.

3.5.2 Compression α

The values of the parameter α describing compression in the auditory system are within the realistic range for humans (e.g. Bacon et al., 2003; Plack, 2004; Rosengard et al., 2005). For model simplicity, the compression is assumed to be independent of the stimulus level. This suffices as an explanation for the general trends in the data presented in this study. A more realistic level dependent compression would have been possible but at the cost of a more complex model which could not have been treated analytically. Due to the fact that the different experiments described in the present study were performed with different subjects and at varying overall stimulus levels, adjusting the compressive exponent to the different experiments is a reasonable assumption.

3.5.3 Signal detection criterion D and internal error σ

The criterion for signal detection $D = 1.8$ has been chosen because of the fact that two Gaussian distributions with their means separated by 1.8 standard deviations can be distinguished quite well and because this criterion has been used in other studies as well (e.g. Klump and Nieder, 2001; Langemann and Klump, 2001; Bee et al., 2007). As discussed in Section 3.3.1, this parameter can be scaled by choosing a different value for σ and may therefore be regarded as a constant rather than a free parameter of the model.

The parameter σ for the supposedly constant internal error that the auditory system faces when detecting changes in the mean compressed envelope has been chosen by fitting the model thresholds to the data. The variability of the estimate of the compressed envelope by the auditory system is affected by at least two contributions: on the one hand, the signal statistics yield a time-varying envelope value, and on the other hand, the partly random firing of auditory nerve fibers makes the number of spikes in a certain time interval a random number. In Buschermöhle et al. (2006) it is assumed that the standard deviation of firing rates of small populations of neurons in the starling's auditory forebrain only marginally depends on the level of the signal in CDD experiments (see also Section 2.7). This may be taken as a hint that the biggest contribution to σ is the firing rate variability due to random spiking and not due to the stimulus variability, as the variability of the stimulus envelope in CDD experiments changes with signal level. For simplicity, the parameter σ is regarded as a constant although the firing rate variability of auditory nerve neurons may depend on the overall level of a stimulus. Surely there is also an intersubject variability concerning the exact value of σ . The different psychophysical procedures and stimulus or signal durations used in

the experiments can affect this parameter as well (Section 2.7 demonstrates the influence of signal duration on the firing rate standard deviation). This may be an explanation for the fact that σ needs to be adjusted slightly in order to model the discussed experiments.

3.5.4 *Parameter dependence of the model results*

Even if the model has only three free parameters, its results may depend critically on their specific choice. To get an idea of the influence of a certain parameter set on the model, we discuss the results for an example stimulus setup for the range of parameters used in this chapter (see Fig. 3.9 and Table 3.2). The example stimulus setup is the first frequency spacing CMR experiment discussed in Sect. 3.4.2. The dependence of thresholds on the compression α for given σ and γ (Fig. 3.9 B) is easily understood: The larger α , the quicker do the mean compressed envelopes increase with signal level and the lower is the signal level at which the signal is detected. Also the general dependence of thresholds on σ (Fig. 3.9 C) is easy to explain: larger values of σ cause the d' -curves to reach the detection criterion D at higher signal levels, which means that the signal is harder to detect and the thresholds increase to higher levels. The influence of the filter bandwidth γ on the model thresholds (Fig. 3.9 D) is to increase the frequency region in which the noise band centered on the signal frequency and the flanking bands can interact to influence the mean compressed envelopes. In the example shown here, a broader filter (larger γ) leads to a broader region of increased thresholds around $\Delta f = 0$ Hz.

One can generally say that by increasing α or by decreasing σ the absolute value of the thresholds can be adjusted while by reducing γ , the size of the affected frequency region can be reduced.

3.5.5 *Influence of temporal processing*

For determining the mean value of compressed envelopes in the model, the average with respect to infinite time and the whole ensemble of possible stimulus realizations is computed. This means that the duration of intervals in the experiments is not explicitly taken into account. There are two possible ways of extending the model by considering the stimulus duration: on the one hand, the parameter σ may be changed accordingly (larger σ for shorter signal intervals). On the other hand, one can perform Monte Carlo simulations of the experiments and use the realizations of stimuli to compute mean compressed envelope values. These kinds of simulations may also be used for

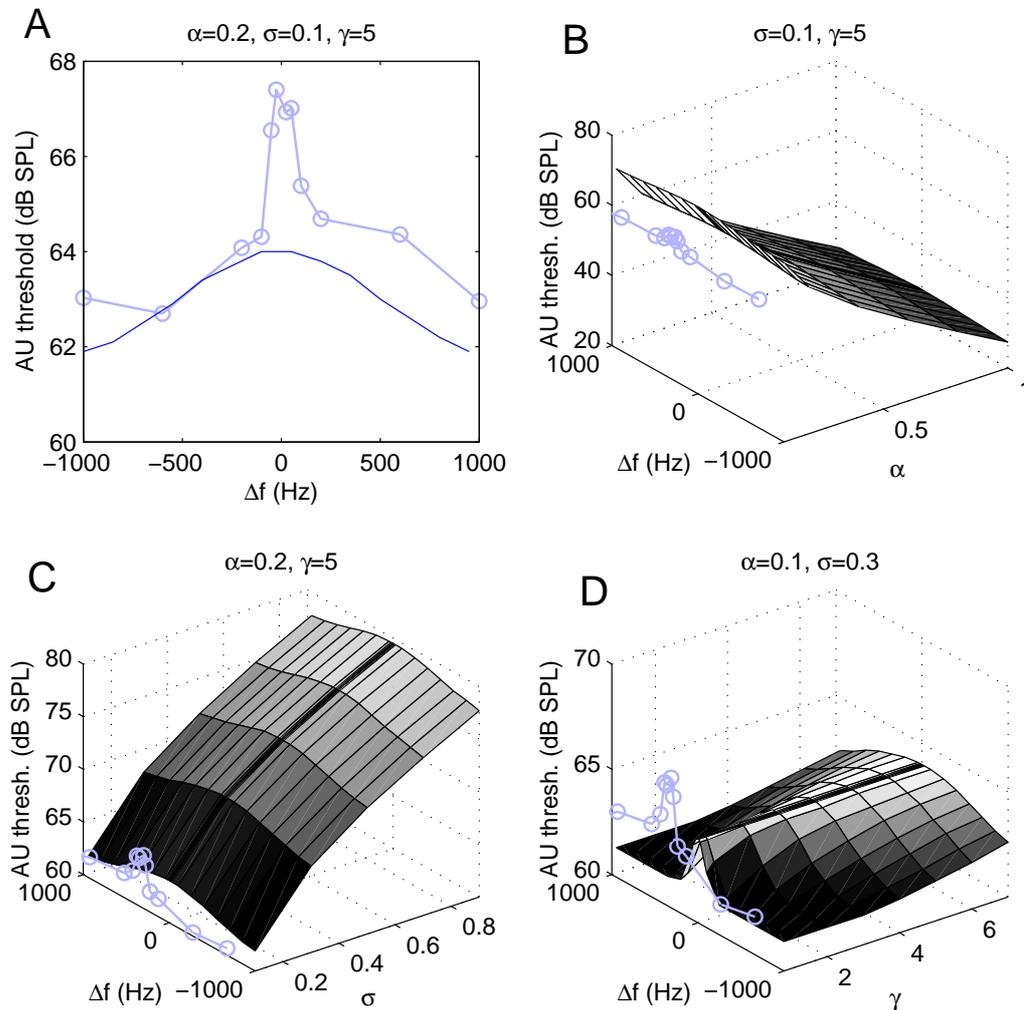


Figure 3.9: Influence of parameter choice on model results for the exemplary case of data from Schooneveldt and Moore (1987). **A:** Replot of the AU signal detection thresholds shown in the top left subplot of Fig.3.5. **B, C, D:** Dependence of the AU threshold on α , σ , and γ , respectively. Each surface plot shows the thresholds in the AU condition predicted by the model while the light colored connected symbols show the experimental data.

determining signal detection thresholds in cases where the stimulus statistics is not calculated as easily as in the cases discussed here.

A further aspect that indicates a possible oversimplification in the proposed model is that for human listeners it is possible to perceive the temporal fine structure of the stimuli. This is not included in the model, as the model averages over time. Different perceptual impressions of maskers alone and maskers plus signal in the model are only possible for one stimulus attribute,

which might correspond to the overall loudness. But surely, subjects do not only perceive loudness differences in stimuli with and without signal. Therefore, the model might be improved by taking into account temporal aspects of the stimuli (such as it is e.g. assumed in Verhey et al., 1999).

3.6 CONCLUSIONS

In this chapter, a very simple within-channel model essentially based on peripheral processing is introduced that is able to reproduce many general aspects of psychophysical CDD and CMR experiments while being derived from physiological investigations. Despite the fact that the model is very parsimonious (only three free parameters) and involves nearly no higher level processing, many aspects of experimental results can be reproduced reasonably well. The model gives insights into the causes of psychophysical thresholds and their dependence on different experimental parameters. One notable feature of the model is the possibility of performing analytical calculations which gives a basic understanding of the importance of the parameters and how they may change the model's behavior. The proposed model constitutes a unifying approach to CDD and CMR and may be instrumental in developing more sophisticated numerical models for experiments with comodulated stimuli.

3.7 APPENDIX: ENVELOPE STATISTICS WITH LEVELS

Denoting the overall level of noise band k by L_k and the overall level of the signal sinusoid for CMR by L_b , Eqns. 3.3–3.5 can be written in terms of levels instead of sinus amplitudes. The RMS-value of a sum of $2N + 1$ sinusoids of amplitude a with independent random phases is $a\sqrt{(2N + 1)/2}$. Setting the reference sound pressure to an RMS value of p_0 (for dB SPL, $p_0 = 20\mu\text{Pa}$), the level of the mentioned sum of sinusoids is $L = 20 \log_{10}(\frac{a\sqrt{(2N+1)/2}}{p_0})$. This expression can be solved for a to give

$$a = p_0 10^{\frac{L}{20}} \sqrt{\frac{2}{2N + 1}}. \quad (3.11)$$

Inserting Eq. 3.11 into Eqns. 3.3–3.5 and remembering that $L_b = 20 \log_{10}(\frac{b}{\sqrt{2}p_0})$ when normalizing to an RMS value of p_0 , one gets:

$$\mu_{\text{AU}}^2 = 2p_0^2 d_0^2 10^{\frac{L_b}{10}} + 2p_0^2 \sum_{k=0}^K 10^{\frac{L_k}{10}} \cdot d_k^2 \quad (3.12)$$

$$\begin{aligned} \mu_{\text{AC}}^2 &= 2p_0^2 d_0^2 10^{\frac{L_b}{10}} + 2p_0^2 \sum_{k,k'=0}^K 10^{\frac{L_k}{20}} 10^{\frac{L_{k'}}{20}} \cdot \\ &\quad d_k d_{k'} \cos(2\pi|f_k - f_{k'}|t) \end{aligned} \quad (3.13)$$

$$\begin{aligned} \mu_{\text{CU}}^2 &= 2p_0^2 d_0^2 10^{\frac{L_b}{10}} + 2p_0^2 d_0^2 10^{\frac{L_0}{10}} + 2p_0^2 \sum_{k,k'=1}^K 10^{\frac{L_k}{20}} 10^{\frac{L_{k'}}{20}} \cdot \\ &\quad d_k d_{k'} \cos(2\pi|f_k - f_{k'}|t) \end{aligned} \quad (3.14)$$

Eqns. 3.12–3.14 are independent of the bandwidth of the noise bands and the frequency spacing $\Delta\nu$ of the component sinusoids. They rely on the assumptions that all phases of the summed sinusoids within one band are independent and random and that the number of added sinusoids is large (i.e. $N \gg 1$).

The equations do not hold if two identical bands are superimposed at the same frequency (this happens for band spacing experiments in the AC condition for $\Delta f = 0$ Hz). In that case, the two superimposed bands with their sine amplitudes a_1 and a_2 can be viewed as one band with the sine amplitude $a_1 + a_2$ resulting in an RMS-value of $(a_1 + a_2)\sqrt{(2N + 1)/2}$, which changes Eqns. 3.12–3.14 accordingly.

CHAPTER 4

ENHANCED SIGNAL DETECTABILITY IN COMODULATED NOISE INTRODUCED BY COMPRESSION¹

4.1 ABSTRACT

Many examples of natural noise show common amplitude modulations at different frequency regions. This kind of noise has been termed comodulated noise and is used predominantly in hearing research, where an enhanced detectability of pure tones and narrow noise bands in comodulated noise compared to unmodulated noise is well known as the CMR or CDD effects, respectively. Here it is shown that only one signal processing step, a compressive nonlinearity motivated by the peripheral auditory system, is sufficient to explain a considerable contribution to these effects. Using an analytical approach, the influence of compression on the detectability of periodic and narrow band signals in the presence of unmodulated and comodulated noise is investigated. This theoretical treatment allows for identifying the mechanism leading to improved signal detection. The compressive nonlinearity constitutes an adaptive gain which selectively boosts a stimulus during time spans of inherently increased signal-to-noise ratio and attenuates it during time spans dominated by noise. On average, these time spans are more pronounced in stimuli with comodulated noise than with unmodulated noise, thus giving rise to the observed CMR and CDD effects.

4.2 INTRODUCTION

Signal detection in noisy environments is performed by all living organisms as well as by many technical devices. In many cases, the background noise in these situations may be assumed to be Gaussian white noise. On other occasions, the noise can be described as colored noise. However, a typical

¹A modified version of this chapter has been submitted (Buschermöhle, Feudel, and Freund, 2007a). The modifications are subtle and comprise only changes of individual words.

property of many naturally generated kinds of noise is that they have common amplitude modulations at various frequency regions (Nelken et al., 1999). Such a kind of noise has been termed *comodulated noise* (see e.g. Hall et al., 1984).

For sine tones as signals, it is well known from hearing research that signal detection is facilitated in comodulated noise compared to unmodulated noise for humans and other animals (e.g. Hall et al., 1984; Langemann and Klump, 2001). The effect that signal detection thresholds for sine tones are lower in comodulated noise than in uncorrelated noise has been termed *comodulation masking release* (CMR). In the real world, most signals that need to be detected are no pure tones but have a certain spectral width, as for example animal calls. Also for narrowband noise signals an improved signal detectability in the presence of comodulated noise compared to unmodulated noise is known from hearing research if the signal's inherent amplitude modulations are different from those of the comodulated noise. The resulting threshold difference has been called the *comodulation detection difference* (CDD). CDDs have been found for humans (e.g. McFadden, 1987; Borrill and Moore, 2002) as well as for animals (Jensen, 2007; Langemann and Klump, 2007).

Different models which treat either CDD or CMR have been published. Most of them are rather complex and aim at mimicking the signal processing steps of the auditory system accurately (Verhey et al., 1999; Borrill and Moore, 2002; Meddis et al., 2002). Also simpler models have been proposed that still follow the processing steps of the auditory pathway (e.g. Buscher-möhle et al., 2006). Many of these models involve compression, which is a signal processing step that is present in the healthy peripheral auditory system (see e.g. Carlyon and Jaysurya Datta, 1997; Oxenham and Bacon, 2003, and references therein). Here, we demonstrate that a compressive nonlinearity is a sufficient minimal condition for the CMR and CDD effects and develop an analytically tractable model exhibiting both effects within the same theoretical framework. Using this model, we quantify the possible contribution of peripheral compression to the observed threshold differences in comodulation experiments.

For carrying out our investigations, we will first introduce a suitable definition of comodulated noise and compare its properties to the properties of Gaussian white noise. Subsequently, we will derive approximations for a modified version of the signal-to-noise ratio (SNR) for stimuli consisting of signals in comodulated and unmodulated noise after a simple nonlinear transformation (compression). These analytical expressions will help to identify the mechanisms leading to the threshold differences. Finally, we will discuss the implications of our findings for experiments and their interpretation in

hearing research. Our conclusions may be generalized to other systems that have to perform similar tasks on acoustic, electromagnetic, seismic or other signals.

4.3 PROPERTIES OF COMODULATED NOISE

In spite of the simple sounding definition of comodulated noise as being noise with common amplitude modulations at different frequency regions, there have been many different ways of generating comodulated noise in the literature. In some studies, comodulated noise was generated by adding up several narrow bands of noise at different center frequencies that shared the same temporal envelope (for example Schooneveldt and Moore, 1987; Ernst and Verhey, 2006). In the present chapter, we follow the approach of other studies (e.g. Hall et al., 1984; Verhey et al., 1999), which define comodulated noise by multiplying two independent time signals, a broadband noise $\eta(t)$ and a comodulator $\rho(t)$:

$$\chi(t) = \rho(t) \cdot \eta(t) \quad (4.1)$$

Different authors selected various choices for $\rho(t)$ and $\eta(t)$. For example, Verhey et al. (1999) used a broadband Gaussian noise (cutoff frequency 10 kHz) as $\eta(t)$ and a narrowband lowpass noise (varying bandwidths on the order of 100 Hz) as the comodulator $\rho(t)$. On the other hand, Bacon and Lee (1997) used a sine tone as the comodulator $\rho(t)$ for a noise band $\eta(t)$ that was centered at a certain frequency. Here, we use Gaussian white noise (cutoff at the Nyquist frequency, $F_\eta = f_s/2$, where f_s is the sampling rate) with a standard deviation of σ_η as $\eta(t)$ and Gaussian lowpass noise (cutoff frequency F_ρ) as $\rho(t)$. Since we choose $\langle \rho \rangle = \langle \eta \rangle = 0$, the standard deviation of $\chi(t)$ is $\sigma_\chi = \sigma_\rho \sigma_\eta$. In the following, we will use unit standard deviation for $\eta(t)$ and $\rho(t)$, and therefore $\sigma_\chi = \sigma_\eta = \sigma_\rho = 1$. Sample traces of unmodulated and comodulated noise are compared in Fig. 4.1. As long as $F_\rho \ll F_\eta$, the comodulation may also be viewed as randomly time varying noise intensity.

The power spectra of comodulated and unmodulated noise as defined above are indistinguishable (cf. Fig. 4.1 C, D). This can be seen by considering that the power spectrum of white noise is flat up to the Nyquist frequency in simulations (or extending to infinity in the idealized case) and that the power spectrum of the comodulator is a box extending up to F_ρ . The Fourier transform of a product of time signals is the convolution of the Fourier transforms of the individual signals. Therefore, the power spectrum of the comodulated noise $\chi(t)$ is flat up to the frequency $F_\eta - F_\rho$ and then falls off linearly to zero at $F_\eta + F_\rho$ (this holds if $F_\rho \ll F_\eta$ and $F_\eta + F_\rho \leq f_s/2$). If $F_\eta = f_s/2$, then due to the aliasing effect (e.g. Priestley, 1983) the nu-

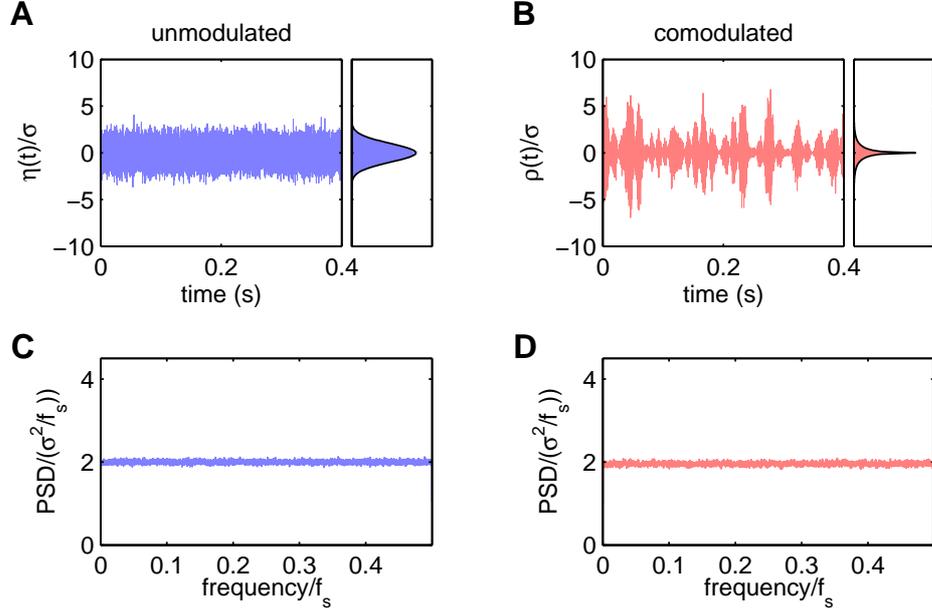


Figure 4.1: Top row: sample traces of unmodulated (**A**) and comodulated (**B**) noise with corresponding amplitude distributions (solid lines in the distribution plots mark theoretical distributions: Normal distribution in **A** and normal product distribution in **B**, see Eq. 4.2). Bottom row: Power spectral densities estimated from time series for unmodulated (**C**) and comodulated (**D**) noise. Parameters: $f_s = 44.1$ kHz, $F_\eta = f_s/2$ $F_\rho = 50$ Hz.

merically obtained spectrum of the comodulated noise $\chi(t)$ cannot be distinguished from a flat spectrum of white noise $\eta(t)$. This also holds true for the idealized case of white noise that has an infinite frequency content. Following from the Wiener-Khinchin theorem (stating that the Fourier transform of the power spectrum of a stationary stochastic process is the autocorrelation function), the autocorrelation function of $\chi(t)$ is δ -peaked at zero time lag and shows no first order correlations in the comodulated noise. The clear temporal structure of comodulated noise therefore must stem from higher order correlations.

As the two noises used for generating $\chi(t)$ are Gaussian, the amplitude distribution of χ is a normal product distribution (see Weisstein, 2002):

$$p_\chi(\chi) = \frac{K_0\left(\frac{|\chi|}{\sigma_\rho\sigma_\eta}\right)}{\pi\sigma_\rho\sigma_\eta}, \quad (4.2)$$

where $K_0(z)$ is a 0th order modified Bessel function of the second kind. The amplitude distribution of the unmodulated Gaussian white noise η obviously is a normal distribution.

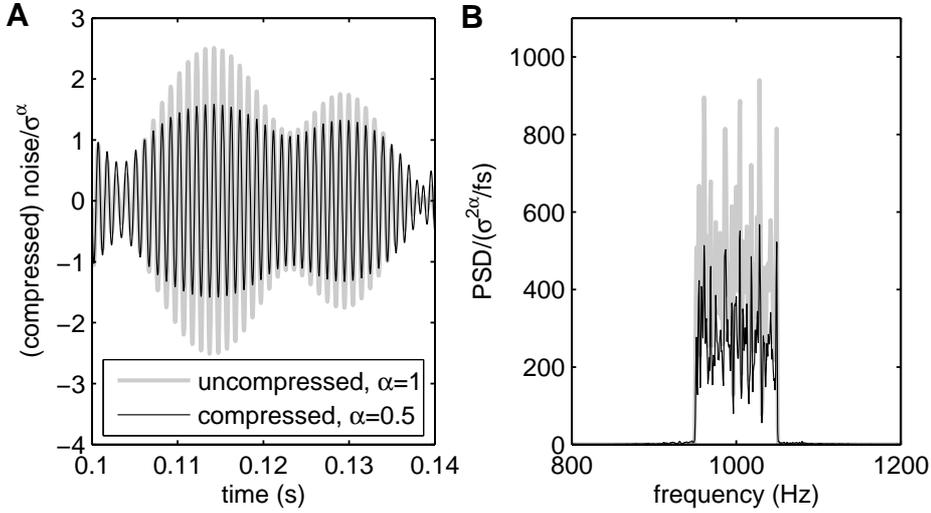


Figure 4.2: The effect of compression on the waveform of a time signal (**A**) and on its power spectrum (**B**) for a 100 Hz wide noise band centered at 1 kHz. There are no power contributions at combination frequencies like 2 kHz or 3 kHz. Data in both panels is plotted such that the plotted quantities are unitless.

The signal used in the present chapter will be denoted as $s(t)$. In the CMR case, it is a sinusoid at the signal frequency f_0 with random initial phase ϕ_0 :

$$s(t) = \cos(2\pi f_0 t + \phi_0). \quad (4.3)$$

In the CDD case, the signal is a bandpass limited noise with corner frequencies $f_0 - \frac{f_{BW}}{2}$ and $f_0 + \frac{f_{BW}}{2}$ which we choose to have unit standard deviation. The complete stimulus (noise masker and sinusoidal or bandpass limited signal) in the comodulated noise condition may be written in the form

$$\gamma_{co}(t) = as(t) + \sigma\chi(t), \quad (4.4)$$

The stimulus in the unmodulated condition is

$$\gamma_{un}(t) = as(t) + \sigma\eta(t). \quad (4.5)$$

In both cases, the masking noise (impairing signal detection) has an intensity σ^2 , while the intensity of the signal is proportional to a^2 . The total intensity of the stimulus is determined by the combination of a and σ .

4.4 ENVELOPE COMPRESSION

One of the most important features of signal processing in the inner ear apart from frequency filtering is the incorporation of a compressive nonlinearity (see

Bacon et al., 2003; Oxenham and Bacon, 2003; Buchholz and Mourjopoulos, 2004). This means that the intensity of the processed stimulus does not increase linearly with the intensity of the input stimulus but rather increases with a reduced slope of less than one dB per dB. There have been many approaches for modeling compression (see e.g. Harte et al., 2005). A possible way for implementing a compressive nonlinearity without introducing considerable frequency distortions has been described in Bernstein et al. (1999) and van der Heijden (2005) as envelope compression or automatic gain control. We use this approach here because it turns out to be analytically tractable. As discussed in Sect. 4.5.4 and 4.5.5, the general results of the present chapter do also occur for other implementations of compression. Envelope compression is implemented by splitting the stimulus into its instantaneous amplitude (the Hilbert envelope) and its instantaneous phase:

$$\gamma(t) = R(t) \cdot \cos(\Phi(t)), \quad (4.6)$$

where $R(t)$ and $\Phi(t)$ are defined via the analytic signal $\mathcal{A}(t)$ (Gabor, 1946) and the Hilbert transform \mathcal{H} (e.g. Hartmann, 1998). The analytic signal of a stimulus $\gamma(t)$ is defined as

$$\mathcal{A}(t) = |\mathcal{A}(t)|e^{i\Phi(t)} = R(t)e^{i\Phi(t)} = \gamma(t) + i\mathcal{H}[\gamma(t)], \quad (4.7)$$

and the Hilbert transform $\mathcal{H}[\gamma(t)]$ of $\gamma(t)$ is given by the integral

$$\mathcal{H}[\gamma(t)] = \frac{1}{\pi} \text{P.V.} \int_{-\infty}^{\infty} \frac{\gamma(t')}{t-t'} dt', \quad (4.8)$$

where the integral is taken in the sense of the Cauchy principal value (P.V.). Now, the stimulus can be compressed using the compressive exponent $0 < \alpha < 1$ by the simple relation

$$\begin{aligned} \hat{\gamma}(t) &= R^\alpha(t) \cos(\Phi(t)) = R^{\alpha-1}(t)R(t) \cos(\Phi(t)) \\ &= R^{\alpha-1}(t)\gamma(t), \end{aligned} \quad (4.9)$$

which introduces only modest frequency distortions due to the fact that the stimulus fine structure is largely contained in the instantaneous phase and not in the instantaneous amplitude being the only compressed part of the stimulus. The influence of the compressive nonlinearity on a stimulus can be most easily understood for a pure sine wave of amplitude a . In that case, the stimulus envelope is a constant, and the compressed stimulus simply remains a sinusoid with amplitude a^α (note that the term *compression* does not imply that the amplitude is reduced but that the growth of the compressed stimulus

with increasing a is reduced compared to the uncompressed stimulus). As an example, the compressed and uncompressed version of a narrow band noise stimulus are depicted in Fig. 4.2 A. The minor effect (on a linear scale) of envelope compression on the spectral shape of the stimulus can be seen in Fig. 4.2 B.

4.4.1 Influence of compression on the standard deviation of unmodulated noise

Without changing the temporal waveforms strongly, the compression has an important effect on the intensity (i.e. variance) of the stimulus. In the following, we will derive an expression for the standard deviation of compressed unmodulated Gaussian white noise which originally had a standard deviation σ .

We denote the original Gaussian white noise by $\eta(t)$. Without compression (i.e. $\alpha = 1$), $\eta(t)$ at any point in time may be regarded as a random variable

$$z = \Re(re^{i\phi}), \quad (4.10)$$

where r is Rayleigh-distributed according to

$$p_r(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (4.11)$$

and ϕ is uniformly distributed,

$$p_\phi(\phi) = \frac{1}{2\pi} \text{ for } \phi \in [-\pi; \pi[. \quad (4.12)$$

The compressed version of $\eta(t)$ may be written as

$$Z = \Re(r^\alpha e^{i\phi}) = r^\alpha \cos \phi. \quad (4.13)$$

If we introduce the new random variable $y = \cos \phi$, then we have

$$p_y(y) = \frac{1}{\pi \sqrt{1 - y^2}}. \quad (4.14)$$

Introducing a further random variable $x = r^\alpha$, we can write

$$\begin{aligned} p_x(x) &= p_r(r) \cdot \left| \frac{dr}{dx} \right| \\ &= \frac{1}{\sigma^2 \alpha} x^{\frac{2}{\alpha} - 1} \exp\left(-\frac{x^{\frac{2}{\alpha}}}{2\sigma^2}\right) \text{ for } x \in [0; \infty[\end{aligned} \quad (4.15)$$

The distribution of the random variable $Z = xy$ can be calculated as (see e.g. Papoulis and Unnikrishna Pillai, 2002):

$$f_Z(Z) = \int_{y=-1}^1 dy \frac{1}{|y|} \frac{1}{\sigma^2 \alpha \pi \sqrt{1-y^2}} \left(\frac{Z}{y}\right)^{\frac{2}{\alpha}-1} \cdot \exp\left(-\frac{\left(\frac{Z}{y}\right)^{2/\alpha}}{2\sigma^2}\right) \Theta\left(\frac{Z}{y}\right) \quad (4.16)$$

Here, $\Theta(\cdot)$ denotes the Heaviside step-function. As the expectation value of Z is zero, the variance of Z may be calculated by determining $E(Z^2)$. This results in the following formula for the standard deviation of compressed Gaussian white noise:

$$\text{std}(\eta^\alpha) = \sqrt{2^{\alpha-1} \Gamma(\alpha+1)} \sigma^\alpha \quad (4.17)$$

In the above equation, η^α is used as a symbol for the compressed unmodulated noise (cf. Eq. 4.13), not for $\eta(t)^\alpha$.

4.4.2 Influence of compression on the standard deviation of comodulated noise

The standard deviation of compressed comodulated noise can be approximated analytically. We write down the uncompressed (but comodulated) noise $\chi(t)$ in the same way as in Eq. 4.10. This time we need to make an assumption about the distribution of r . As long as $F_\rho \ll F_\eta$, an exponential distribution is a reasonable approximation to this distribution (see Appendix 4.7 and Fig. 4.6). Therefore,

$$p_r(r) = \frac{1}{\sigma} e^{-\frac{r}{\sigma}}. \quad (4.18)$$

The phase ϕ may again be assumed to be uniformly distributed. Using the same argumentation as before, one finds the standard deviation of the compressed comodulated noise to be

$$\text{std}(\chi^\alpha) = \sqrt{\frac{1}{2} \Gamma(2\alpha+1)} \sigma^\alpha \quad (4.19)$$

The predictions of Eqns. 4.17 and 4.19 are compared to simulations in Fig. 4.3, where one can see that the standard deviation of compressed noise ($\alpha < 1$) increases more slowly than that of uncompressed noise ($\alpha = 1$). For comparison, we also show the case of expanded noise ($\alpha > 1$). This case is physiologically not realistic and will therefore not be considered in the following.

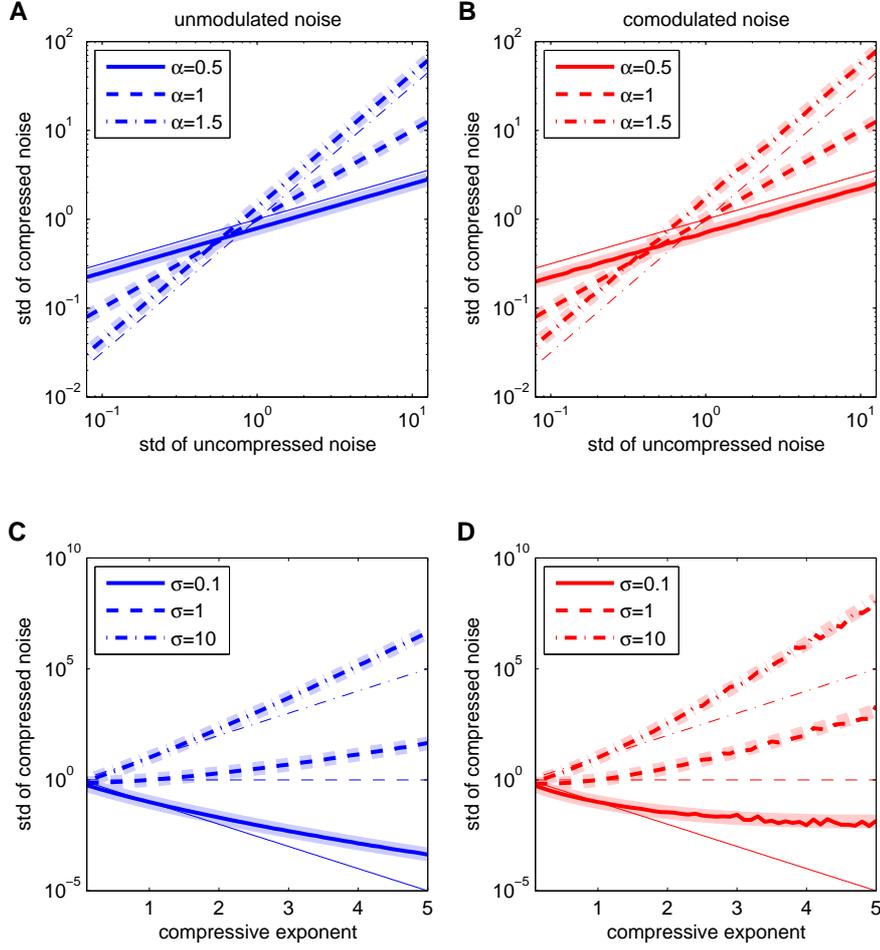


Figure 4.3: Effect of compression on the intensity of unmodulated and comodulated noise. In all panels, the thin lines show the functions σ^α while the thick dark lines show simulations and the very thick light colored lines show the predictions by Eqns. 4.17 and 4.19. **A, B:** standard deviation of compressed noise depending on standard deviation of uncompressed noise for unmodulated and comodulated noise, respectively, for three different values of α , the compressive exponent. **C, D:** standard deviation of compressed noise depending on α for unmodulated and comodulated noise, respectively, for three different values of initial standard deviations.

4.5 SIGNAL DETECTION IN (UN)COMPRESSED NOISE

For quantifying the detectability of a sinusoidal signal in noise (the CMR case), we employ a variant of the signal-to-noise ratio (SNR) that is also used in the literature on stochastic resonance (McNamara and Wiesenfeld, 1989). There, the SNR is defined as the total amount of power contained in a frequency region of width Δf centered on the signal frequency f_0 divided

by the power in this region without the signal contribution:

$$\text{SNR} = \frac{N\Delta f + S}{N\Delta f}. \quad (4.20)$$

When writing this equation, we assume that the spectrum within a region of size Δf around the signal frequency can be approximated as the sum of a flat noise contribution N and a δ -peaked signal contribution $S\delta(f - f_0)$.

With this definition, one could rather speak of the signal-plus-noise-to-noise ratio. We use this measure because it has two advantages (see McNamara and Wiesenfeld, 1989): First, it can be more easily estimated in numerical simulations (and experiments) with small signal amplitudes than the usual definition of the SNR because the numerator can be estimated by integrating the power spectral density over a narrow region around the signal frequency without having to subtract the noise contribution. Second, for vanishing signal, its value becomes 1.0 (or 0 dB, respectively) instead of having a singularity on a logarithmic scale. Thus, our definition of the SNR is presumably more directly comparable to measures the auditory system may use for signal detection.

The SNR for the case of a narrowband noise signal of bandwidth f_{BW} is analogously defined as

$$\text{SNR} = \frac{Nf_{BW} + Sf_{BW}}{Nf_{BW}}. \quad (4.21)$$

where the spectrum is assumed to consist of a flat noise part and a box-shaped signal part.

We assume that the signal can be detected as soon as the SNR exceeds a certain value. The signal level at which this value is reached is the signal detection threshold. According to the experimental evidence described earlier, we expect that signal detection thresholds are lower in comodulated noise than in unmodulated noise.

We will now derive analytical approximations for the SNR of compressed stimuli for several stimulus configurations. For this purpose, we first need to approximate the compressed stimulus (for small signal amplitudes linear in a , for large signal amplitudes linear in σ). Then we will calculate the autocorrelation function of this approximated compressed stimulus. Finally, we determine the power spectral density as the Fourier transform of the autocorrelation function, following the Wiener-Khinchin theorem. From the spectrum, the SNR can be determined in a straightforward manner. For our calculations, there are four different cases to be considered:

- CDD, small signal (i.e. $a < \sigma$),

- CDD, large signal (i.e. $a > \sigma$),
- CMR, small signal (i.e. $a < \sigma$),
- CMR, large signal (i.e. $a > \sigma$).

The necessary calculations are carried out in Appendix 4.8. We will state the results in the following section.

4.5.1 SNR for narrowband noise signals (CDD)

The CDD stimulus consists of the broadband un- or comodulated noise masker

$$m(t) = \begin{cases} \eta(t) & \text{for unmodulated noise} \\ \chi(t) = \rho(t)\eta(t) & \text{for comodulated noise} \end{cases} \quad (4.22)$$

and a narrowband noise signal $s(t)$, yielding

$$\gamma(t) = \sigma m(t) + a s(t) = \Re(r_m e^{i\phi_m} + r_s e^{i\phi_s}). \quad (4.23)$$

In the second step of this equation we made use of the notation introduced in Sect. 4.4 by splitting the masker and the signal into their instantaneous envelope and fine structure and writing the stimulus as the real part of the analytic signal, keeping in mind that r_m , ϕ_m , r_s , and ϕ_s are actually $r_m(t)$, $\phi_m(t)$, $r_s(t)$, and $\phi_s(t)$.

With this notation, the calculations in Appendix 4.8 result in the following expressions for the SNR. The SNR for small signals is

$$\text{SNR}_{\text{CDD,small}} \approx 1 + \frac{\alpha a^2 \langle r_m^{\alpha-1} \rangle^2}{\left[\frac{1}{2} \langle r_m^{2\alpha} \rangle + \frac{1}{2} (\alpha^2 + 1) a^2 \langle r_m^{2\alpha-2} \rangle - \alpha a^2 \langle r_m^{\alpha-1} \rangle^2 \right] \frac{f_{BW}}{F_\eta}}, \quad (4.24)$$

and the SNR for large signals is

$$\text{SNR}_{\text{CDD,large}} \approx 1 + \frac{2^{\alpha-1} \Gamma(\alpha + 1) a^{2\alpha}}{\frac{1}{4} (\alpha^2 + 1) \langle r_m^2 \rangle 2^{\alpha-1} a^{2\alpha-2} \Gamma(\alpha) \frac{f_{BW}}{F_\eta}}. \quad (4.25)$$

These expressions hold for unmodulated as well as for comodulated noise. For unmodulated noise, r_m is Rayleigh-distributed, and the expectation values for any power of r_m can be given analytically as $\langle r_m^z \rangle_{\text{un}} = 2^{z/2} \sigma^z \Gamma(z/2 + 1)$. For comodulated noise, the distribution of r_m can be approximated by an exponential distribution (see Appendix 4.7), giving also analytical expressions for powers of r_m , namely $\langle r_m^z \rangle_{\text{co}} = \sigma^z \Gamma(z + 1)$. This approximation is not very good for strong compression (i.e. small values of α) because of the

fact that the actual distribution of r_m for comodulated noise differs from an exponential distribution for very small values of r_m , and these values gain increasing importance for stronger compression. When comparing numerical simulations with the analytical expressions, we therefore do not obtain the expectation values for r_m for comodulated noise from the exponential distribution but rather estimate them numerically.

4.5.2 SNR for pure tone signals (CMR)

For the CMR case, the signal is $s(t) = \cos(\omega_0 t + \phi_0)$, where ϕ_0 is a random initial phase. The masking noise $m(t)$ stays the same as before, and therefore, we can write the stimulus as

$$\gamma(t) = \sigma m(t) + a s(t) = \Re(r_m e^{i\phi_m} + a e^{i\omega_0 t + i\phi_0}). \quad (4.26)$$

As described in Appendix 4.8, the SNR for small signals can be approximated as

$$\text{SNR}_{\text{CMR,small}} \approx 1 + \frac{\frac{1}{8} a^2 (\alpha + 1)^2 \langle r_m^{\alpha-1} \rangle^2}{\left[\frac{1}{2} \langle r_m^{2\alpha} \rangle + \frac{1}{4} a^2 (\alpha^2 + 1) \langle r_m^{2\alpha-2} \rangle - \frac{1}{8} a^2 (\alpha + 1)^2 \langle r_m^{\alpha-1} \rangle^2 \right] \frac{f_{BW}}{F_\eta}}, \quad (4.27)$$

and the SNR for large signals is given by

$$\text{SNR}_{\text{CMR,large}} \approx 1 + \frac{\frac{1}{2} a^{2\alpha}}{\frac{1}{4} (\alpha^2 + 1) \langle r_m^2 \rangle a^{2\alpha-2} \frac{f_{BW}}{F_\eta}}. \quad (4.28)$$

Again, the expectation values for r_m may be expressed analytically by assuming a Rayleigh-distribution in the unmodulated noise case and an exponential distribution in the comodulated noise case.

4.5.3 Improved signal detectability in comodulated noise after compression

Signal-to-noise ratios for different amounts of compression and for the two possible kinds of signal are plotted in Fig. 4.4. For the case of a pure sinusoid as signal (that is, for the CMR case), the top row of that figure shows the numerically obtained SNRs in the compressed stimuli as markers. The numerics are compared to the analytical expressions derived from Eqns. 4.27 and 4.28.

The bottom row of Fig. 4.4 compares the numerical data and the analytical expressions derived from Eqns. 4.24 and 4.25 for a narrow band of noise as signal (the CDD case). Although in Fig. 4.4, the signal levels are given

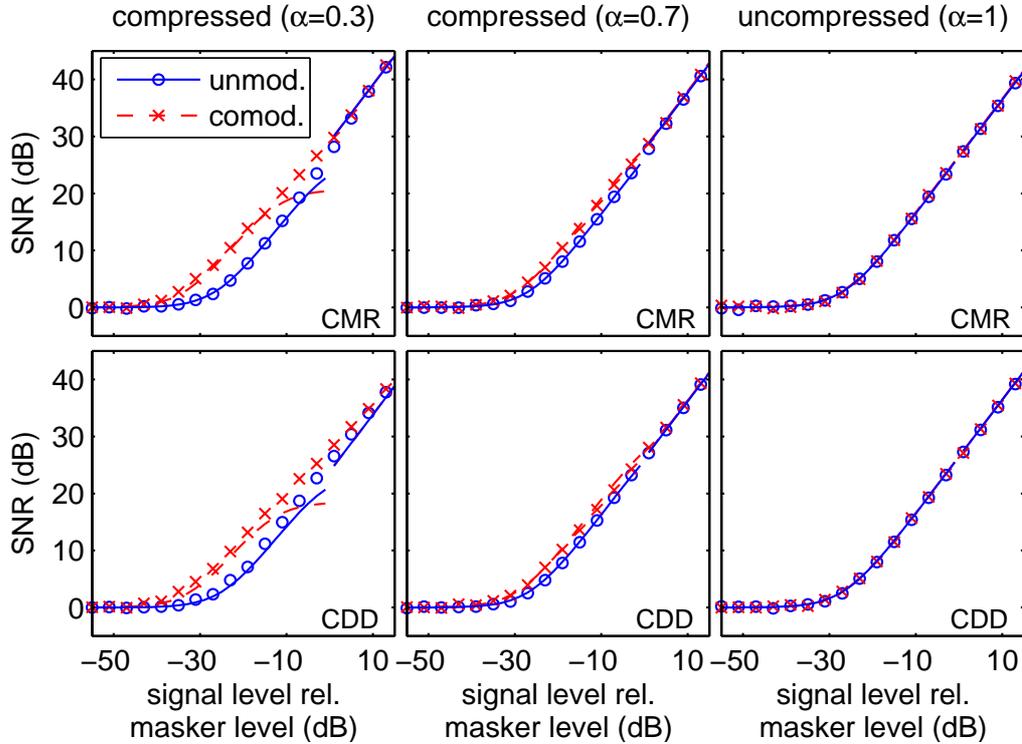


Figure 4.4: Signal-to-noise ratio (SNR, see Eqns. 4.20 and 4.21) in relation to the signal level for compressed (first two columns) and uncompressed (last column) noise. Top rows show the case for a pure tone signal (the CMR paradigm). Bottom rows show the case for a narrow band noise signal (the CDD paradigm). Lines denote theoretical approximations according to Eqns. 4.24, 4.25, 4.27, and 4.28. Symbols show results from numerical simulations. The stronger the compression, the larger are the differences for signal detection between comodulated and unmodulated noise. Parameters: $f_0 = 2.0$ kHz, $f_{BW} = F_\rho = 50$ Hz, $f_s = 44.1$ kHz, masker level is 60 dB SPL. Signal levels are measured relative to the masker level.

relative to the masker level, the expressions for the SNR are not functions of a/σ , which means that the figure looks slightly different for other masker levels.

For both cases of CMR and CDD, without compression ($\alpha = 1$), the SNR does not differ for the two kinds of masking noise. This is how the stimuli are constructed. If α is reduced towards zero, the amount of compression increases. At the same time, due to the nonlinear transform of the stimulus, the SNR curves for the different kinds of maskers are not the same anymore. While for small signal levels, the masker dominates the SNR and for large signal levels the signal dominates the SNR, for intermediate signal levels, the SNR is influenced by the signal and the masker at the same time. In

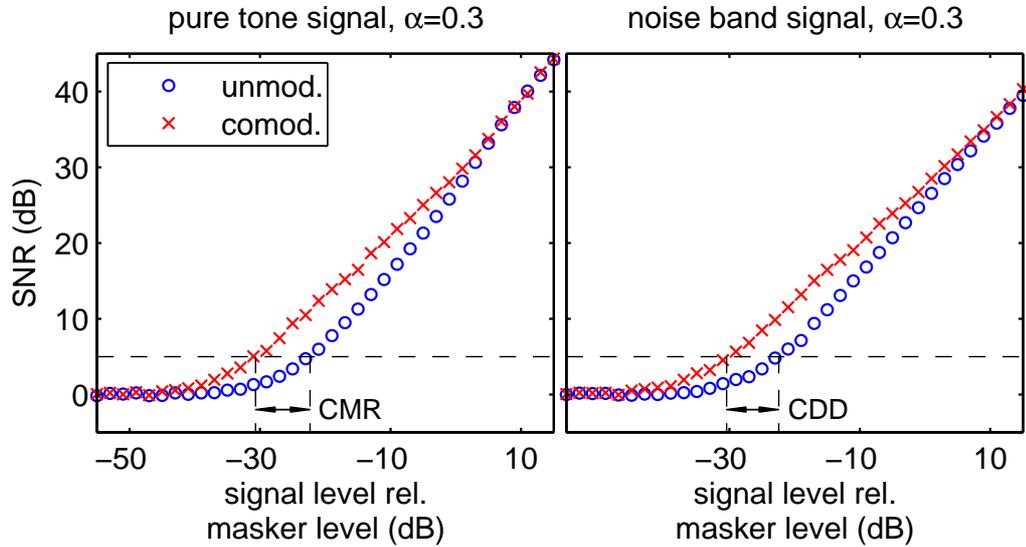


Figure 4.5: Same data as in the leftmost column of Fig. 4.4, but only the simulation results with a better resolution. In both panels, the dashed horizontal line marks a detection criterion of an SNR of 5 dB. The dashed vertical lines indicate the signal detection thresholds according to this criterion. A double arrow indicates the threshold difference, which corresponds to the CMR (comodulation masking release) in the left panel and the CDD (comodulation detection difference) in the right panel. The threshold difference in both cases is roughly 8.5 dB for the CMR case and 8 dB for the CDD case.

this range of signal levels, the SNR for stimuli with comodulated compressed noise raises earlier than that for stimuli with unmodulated compressed noise. The stronger the compression, the larger the difference between both kinds of noise gets.

These results show that a signal may be detected more easily in comodulated noise than in unmodulated noise if the stimulus is compressed. This corresponds to what is known experimentally from psychophysical research. In the experiments, the differences between signal detection thresholds are determined. Signal detection thresholds in our case may be defined as the signal level at which a certain SNR value is reached. Fig. 4.5 again shows the numerically obtained SNR values for a compressive exponent of $\alpha = 0.3$. Compressive exponents close to this value have been used in auditory modeling (e.g. Lopez-Poveda and Meddis, 2001; Sumner et al., 2002) and are found in experiments with humans and other animals (cf. Köppl and Yates, 1999; Plack, 2004). For a signal detection criterion of an SNR of 5 dB, one can determine the signal detection thresholds (indicated by vertical dashed lines in Fig. 4.5) and determine the amount of CMR or CDD that is due only to compression. Although the exact threshold difference depends on the signal

detection criterion, signal detection in comodulated noise can be improved by about 5 dB to 10 dB solely due to a compressive nonlinearity.

4.5.4 Causes of signal detection differences

The fact that we have approximate analytical expressions for the signal and noise contributions to the spectra of the compressed stimuli allows us to get a deeper insight into why compression results in better signal detectability in comodulated noise than in unmodulated noise. We will discuss this for the case of a narrowband noise signal (CDD). The CMR case can be treated analogously.

For the threshold differences, the parts of the SNR-curves with small signal contributions are most important. It turns out that the α -dependent part of the noise contribution to the SNR,

$$N = \left[\frac{1}{2} \langle r_m^{2\alpha} \rangle + \frac{1}{2} (\alpha^2 + 1) a^2 \langle r_m^{2\alpha-2} \rangle - \alpha a^2 \langle r_m^{\alpha-1} \rangle^2 \right] \frac{1}{F_\eta}, \quad (4.29)$$

is smaller than the signal contribution

$$S = \frac{\alpha a^2}{f_{BW}} \langle r_m^{\alpha-1} \rangle^2. \quad (4.30)$$

This observation is true for unmodulated as well as for comodulated noise and is explained by the fact that $f_{BW} \ll F_\eta$. Therefore, the main factor influencing the initial increase in SNR is the signal contribution S . For given compression α and comodulator bandwidth f_{BW} , the only difference in S for unmodulated and comodulated noise can be located in $\langle r_m^{\alpha-1} \rangle^2$. The reason for S_{co} being larger than S_{un} is therefore, that $\langle r_m^{\alpha-1} \rangle_{co} > \langle r_m^{\alpha-1} \rangle_{un}$.

Indeed, if we remember that r_m is Rayleigh-distributed in the unmodulated noise case and approximately exponentially distributed in the comodulated noise case, we can calculate the above expectation values as

$$\langle r_m^{\alpha-1} \rangle_{co} \approx \sigma^{\alpha-1} \Gamma(\alpha) \quad (4.31)$$

and

$$\langle r_m^{\alpha-1} \rangle_{un} = 2^{\frac{\alpha-1}{2}} \sigma^{\alpha-1} \Gamma\left(\frac{\alpha+1}{2}\right). \quad (4.32)$$

The ratio of these two expressions for any $\alpha < 1$ is always larger than one. For the case of no compression ($\alpha = 1$), it is equal to one:

$$\frac{\langle r_m^{\alpha-1} \rangle_{co}}{\langle r_m^{\alpha-1} \rangle_{un}} \approx 2^{\frac{1-\alpha}{2}} \frac{\Gamma(\alpha)}{\Gamma\left(\frac{\alpha+1}{2}\right)} \geq 1. \quad (4.33)$$

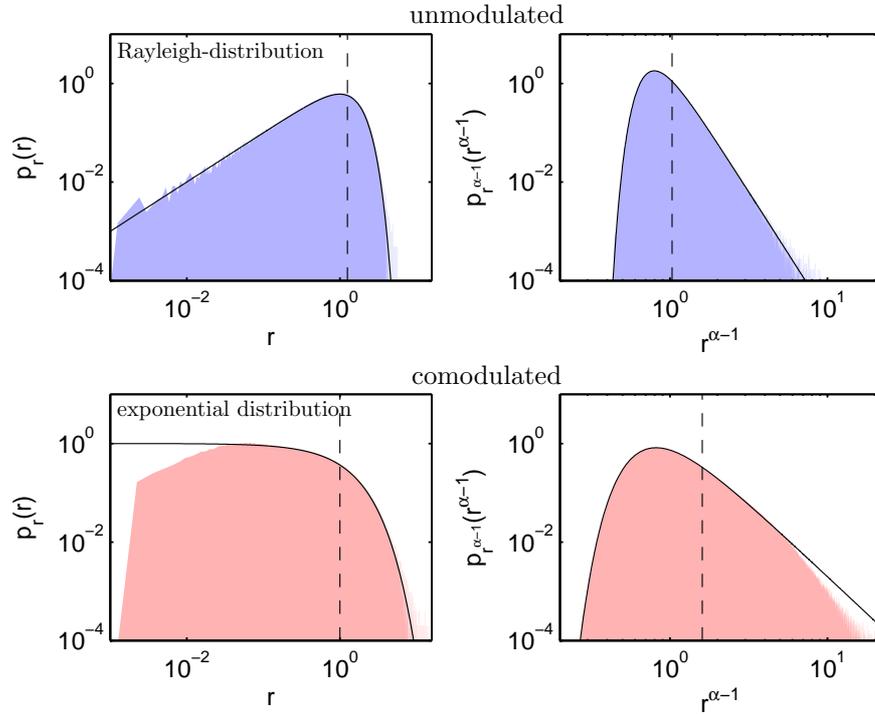


Figure 4.6: Distribution of r_m (left column) and $r_m^{\alpha-1}$ (right column) for unmodulated (top row) and comodulated (bottom row) noise. The patches show distributions obtained from numerical histograms, while the solid lines show analytical expressions (Rayleigh-distribution for r_m in unmodulated noise and exponential distribution for r_m in comodulated noise; the other two distributions can be derived similarly to Eq. 4.15 from these distributions). The numerical mean values are denoted by dashed vertical lines. One can see that the assumption of an exponential distribution for the envelopes of comodulated noise is only approximately correct. Here: $\alpha = 0.5$, $F_p = 50$ Hz.

If α decreases towards zero, the above ratio is monotonically increasing, which corresponds to the numerically found fact that the difference between comodulated and unmodulated signal detection thresholds increases monotonically when reducing α towards zero. For very strong compression, the differences between an exponential distribution and the real distribution of r_m in the comodulated noise case become increasingly important, but the general idea one gets from this analytical consideration is still correct. See Fig. 4.6 for a comparison of the numerically obtained and the analytically calculated distributions of r_m and $r_m^{\alpha-1}$.

From these considerations, one can see why there are differences between compressed comodulated noise and compressed unmodulated noise. The factor $r_m^{\alpha-1}$ is roughly the prefactor to the original stimulus when our form of

envelope compression is applied at small signal levels. For small signal amplitudes, we have:

$$\hat{\gamma}(t) = R(t)^{\alpha-1}\gamma(t) \approx r_m(t)^{\alpha-1}\gamma(t). \quad (4.34)$$

Therefore, the on average larger prefactor to the complete stimulus in the comodulated noise condition when applying the compression is the cause of the SNR differences between the two kinds of noise.

One can also interpret compression as automatic gain control (van der Heijden, 2005): if the stimulus gets temporarily very intense, the compression attenuates it, while a temporarily faint stimulus is amplified. This can be seen intuitively when writing

$$\hat{\gamma}(t) = \frac{1}{R(t)^{1-\alpha}}\gamma(t) \approx \frac{1}{r_m(t)^{1-\alpha}}\gamma(t). \quad (4.35)$$

The gain $1/R(t)^{1-\alpha}$ scales up the stimulus when the noise is small. As comodulated noise has more pronounced periods of low noise compared to unmodulated noise (cf. left column of Fig. 4.6), the complete stimulus (including the signal) is amplified more often in a stimulus consisting of comodulated noise with small signals.

This also shows why other forms of compression (not only instantaneous compression) lead to a similar shift in SNR curves for comodulated and unmodulated noise: The gain factor that introduces the compressive nonlinearity boosts the stimulus specifically at periods of inherently high signal-to-noise ratio, such that the CDD and CMR effects will be found for many different kinds of compression.

4.5.5 Example: instantaneous compression

Our results have been derived for the special case of instantaneous envelope compression, but they can be generalized to other kinds of compression as well. As an example for a different compression scheme, we may look at instantaneous compression introduced by compressing the absolute value of the stimulus and multiplying it with the sign of the stimulus (see e.g. Harte et al., 2005). In this case, the compressed stimulus is given as

$$\hat{\gamma}(t) = |\gamma(t)|^\alpha \text{sign}(\gamma(t)) = |\gamma(t)|^{\alpha-1}\gamma(t). \quad (4.36)$$

For unmodulated noise, the distribution of instantaneous amplitude values is Gaussian, while for our form of comodulated noise, it is a normal product

distribution (see Eq. 4.2). In the case of small signals, one gets for the expectation values of the gain factor $|\gamma|^{\alpha-1}$:

$$\langle |\chi|^{\alpha-1} \rangle_{\text{co}} = \frac{1}{\pi} 2^{\alpha-1} \sigma^{\alpha-1} \Gamma\left(\frac{\alpha}{2}\right)^2 \quad (4.37)$$

and

$$\langle |\eta|^{\alpha-1} \rangle_{\text{un}} = \frac{1}{\sqrt{\pi}} 2^{\frac{\alpha-1}{2}} \sigma^{\alpha-1} \Gamma\left(\frac{\alpha}{2}\right). \quad (4.38)$$

For $\alpha < 1$, one finds $\langle |\chi|^{\alpha-1} \rangle_{\text{co}} > \langle |\eta|^{\alpha-1} \rangle_{\text{un}}$, which therefore also introduces differences in the SNR curves. Other kinds of compression can be analyzed in a similar way.

4.6 DISCUSSION AND CONCLUSIONS

We have shown that considerable contributions to the CMR and CDD effects may already be introduced by a compressive nonlinearity which is known to be present in the peripheral auditory system. This conclusion is based on the assumption that signal detection in the auditory system is carried out by a measure related to the SNR as introduced in the present chapter. By means of analytical approximations for the SNR, it could be shown that the cause for this contribution to an improved signal detectability in comodulated noise compared to unmodulated noise is the adaptive gain of the active system employing the compression which is on average larger for comodulated than for unmodulated background noise.

The order of magnitude of the threshold differences found for the CMR and CDD paradigms is the same as found in psychophysical experiments (cf. Borrill and Moore, 2002; Verhey et al., 2003). Still, the threshold differences found in experiments tend to be larger than those found here indicating that other mechanisms besides compression contribute to the CMR and CDD effects. Furthermore, we have disregarded one of the most important signal processing steps of the auditory system, the frequency filtering, in order to be able to treat the problem analytically. If a realistic auditory filterbank is included in the signal processing pathway (e.g. the one described in Hohmann, 2002), the general result of earlier increasing SNR for signals with comodulated noise compared to unmodulated noise remains the same, but the amount of CMR and CDD is reduced. This also suggests further contributions to both effects from other mechanisms.

Our result that compression plays an important role in comodulation experiments is in agreement with experimental findings for hearing impaired subjects with cochlear hearing loss. Usually, the compression in these subjects is reduced (corresponding to $\alpha \approx 1$, see Oxenham and Bacon, 2003).

Generally, subjects with cochlear hearing loss show smaller CMRs than normal hearing subjects (e.g. Hall et al., 1988; Moore et al., 1993; Hall and Grose, 1994; Grose and Hall, 1996). Further experimental support for our findings can be obtained from the fact that compression is strongest (α closest to zero) for intermediate sound pressure levels (above 30 dB SPL), while it is nearly absent for sound pressure levels below that (see e.g. Harte et al., 2005). Therefore, with the results of the present chapter, one expects a reduction of CMR and CDD values for very low masker levels. Indeed, Borrill and Moore (2002) found a reduced CDD for low masker levels. Moore and Shailer (1991) found that the monotic CMR for masker spectrum levels below 30 dB SPL nearly vanishes. Both of these experimental observations are in accordance with the idea of an important contribution of compression to both effects.

Nevertheless, when comparing the results of the present chapter with experimental data, one has to be aware that our stimuli are only similar to those used in experiments, not exactly the same. Especially for the CDD case, the maskers in experiments usually consist of several narrow noise bands instead of being one broadband noise masker. The main difference is that for CDD experiments, the masker has no contribution at the signal frequency, while in our model, the masker has no spectral notch at the frequencies of the signal band. Still, we believe that our model stimulus can be used to understand the contributions of compression to the CDD effect.

An important finding of the present chapter is that the differences between the SNR curves for unmodulated and comodulated masking noise increase monotonically with decreasing α . This means that the optimal compression for utilizing the differences between the two kinds of noise would be a compression that only keeps the temporal fine structure of the stimulus and completely discards the stimulus envelope ($\alpha = 0$). While this may be the optimal compression for the presently discussed task, other tasks that have to be performed by the auditory system (like loudness perception) may require a more subtle compression. This and mechanical constraints may be reasons why the evolution has not led to extreme compression in the auditory system but rather to compressions that may be described by $\alpha \approx 0.3$.

As comodulated noise does not only occur in the auditory environment but possibly also in various other fields of nature or technology, the results of the present chapter indicate that a compressive nonlinearity could be an advantageous processing step for signal detecting devices.

4.7 APPENDIX: ENVELOPE DISTRIBUTION OF COMODULATED NOISE

According to Eq. 4.1, comodulated noise with a standard deviation of σ may be written as $\sigma\chi(t) = \sigma\rho(t) \cdot \eta(t)$, where ρ and η are Gaussian distributed with unit standard deviation. If $F_\rho \ll F_\eta$, one can interpret the comodulated noise as Gaussian noise with a slowly varying standard deviation $\sigma\rho(t)$. As stated in Eq. 4.11, the envelope distribution of Gaussian noise with standard deviation $x = \sigma\rho$ is given by

$$p(r|x) = \frac{r}{x^2} \exp\left(-\frac{r^2}{2x^2}\right). \quad (4.39)$$

If we denote the Gaussian distribution of x with $p_x(x) = \sqrt{2\pi\sigma^2}^{-1} \exp(-x^2/2\sigma^2)$, then we can determine the envelope distribution of comodulated noise by

$$\begin{aligned} p_r(r) &= \int_{-\infty}^{\infty} p(r|x)p_x(x)dx \\ &= \int_{-\infty}^{\infty} \frac{r}{\sqrt{2\pi}x^2\sigma} \exp\left(-\frac{r^2}{2x^2} - \frac{x^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sigma} \exp\left(-\frac{r}{\sigma}\right). \end{aligned} \quad (4.40)$$

This exponential distribution describes the envelope distribution of comodulated noise reasonably well (cf. Fig. 4.6) as long as the temporal variations of the comodulator are slow compared to the fastest variations of $\eta(t)$ and as long as the comodulator varies fast enough to sample an adequate part of the distribution of ρ within the total duration of the comodulated noise.

4.8 APPENDIX: CALCULATIONS OF THE SNR

Here, we will derive Eqns. 4.24, 4.25, 4.27, and 4.28. The cases of a small amplitude narrow band signal and a large amplitude pure tone signal will be demonstrated in detail, while the cases of a large amplitude narrow band signal and a small amplitude pure tone signal can be treated analogously and will therefore be explained only briefly.

4.8.1 Approximations for small narrowband signals (CDD)

As stated in Eq. 4.23, the CDD stimulus consists of the broadband un- or comodulated noise masker $m(t)$ and a narrowband noise signal $s(t)$:

$$\gamma(t) = \sigma m(t) + as(t) = \Re(r_m e^{i\phi_m} + r_s e^{i\phi_s}). \quad (4.41)$$

The compressed stimulus is

$$\hat{\gamma}(t) = |\mathcal{A}(t)|^{\alpha-1}\gamma(t). \quad (4.42)$$

The absolute value of the analytic signal raised to the power of $\alpha - 1$ can be written as

$$\begin{aligned} |\mathcal{A}(t)|^{\alpha-1} &= |r_m e^{i\phi_m} + r_s e^{i\phi_s}|^{\alpha-1} \\ &= [r_m^2 + r_s^2 + 2r_s r_m \cdot \\ &\quad (\cos \phi_m \cos \phi_s + \sin \phi_m \sin \phi_s)]^{\frac{\alpha-1}{2}} \\ &= r_m^{\alpha-1} \left[1 + \frac{r_s^2 + r_s y}{r_m^2} \right]^{\frac{\alpha-1}{2}}, \end{aligned} \quad (4.43)$$

where we substituted $y = 2r_m(\cos \phi_m \cos \phi_s + \sin \phi_m \sin \phi_s)$. Newton's generalized binomial theorem states that the series

$$(1+z)^r = \sum_{k=0}^{\infty} \binom{r}{k} z^k \quad (4.44)$$

converges for $|z| < 1$. Therefore, for small r_s , we can write

$$\begin{aligned} |\mathcal{A}(t)|^{\alpha-1} &= r_m^{\alpha-1} \sum_{k=0}^{\infty} \binom{\frac{\alpha-1}{2}}{k} \left(\frac{r_s^2 + r_s y}{r_m^2} \right)^k \\ &\approx r_m^{\alpha-1} + r_m^{\alpha-3} r_s \frac{\alpha-1}{2} y \end{aligned} \quad (4.45)$$

This results in an approximation for the compressed stimulus linear in r_s :

$$\begin{aligned} \hat{\gamma}(t) &= \gamma(t) |\mathcal{A}(t)|^{\alpha-1} \\ &\approx r_s r_m^{\alpha-1} \cos \phi_s + r_m^\alpha \cos \phi_m + (\alpha-1) r_m^{\alpha-1} r_s \cdot \\ &\quad \cos \phi_m (\cos \phi_m \cos \phi_s + \sin \phi_m \sin \phi_s) \end{aligned} \quad (4.46)$$

For calculating the spectrum, we first look at the autocorrelation function (the individual expectation values of $\hat{\gamma}$ vanish):

$$C_{\hat{\gamma}\hat{\gamma}}(\tau) = \langle \hat{\gamma}(t)\hat{\gamma}(t-\tau) \rangle_{t,r,\phi} \quad (4.47)$$

After some calculations one arrives at

$$\begin{aligned} C_{\hat{\gamma}\hat{\gamma}}(\tau) &\approx \left[\frac{1}{2} \langle r_m^{2\alpha} \rangle + \frac{1}{4} (\alpha^2 + 1) \langle r_s^2 \rangle \langle r_m^{2\alpha-2} \rangle - \right. \\ &\quad \left. \alpha \langle r_m^{\alpha-1} \rangle^2 \langle s(t)^2 \rangle_t \right] \delta(\tau) + \\ &\quad \alpha \langle r_m^{\alpha-1} \rangle^2 \langle s(t)s(t-\tau) \rangle_t. \end{aligned} \quad (4.48)$$

Here, $\langle s(t)s(t - \tau) \rangle_t$ denotes the autocorrelation function of the uncompressed signal band. We can now calculate the power spectral density of the compressed stimulus as the Fourier transform of the autocorrelation function:

$$S_{\hat{\gamma}\hat{\gamma}}(f)_{\text{CDD,small}} \approx \frac{1}{2}\langle r_m^{2\alpha} \rangle + \frac{1}{4}(\alpha^2 + 1)\langle r_s^2 \rangle \langle r_m^{2\alpha-2} \rangle - \alpha \langle r_m^{\alpha-1} \rangle^2 \langle s(t)^2 \rangle_t + \alpha \langle r_m^{\alpha-1} \rangle^2 S_{ss}(f). \quad (4.49)$$

In this expression, $S_{ss}(f)$ is the power spectrum of the uncompressed signal band alone. The signal band is a bandpass limited Gaussian noise with standard deviation a and bandwidth f_{BW} . Therefore, its power spectral density is a box of height a^2/f_{BW} and width f_{BW} centered on the signal frequency f_0 . For comparing the SNR determined from the analytical expressions to the SNR determined in numerical simulations, a factor $2/f_s = 1/F_\eta$ has to be included in the following expression for the signal-to-noise ratio from the on-sided power spectral density:

$$\text{SNR}_{\text{CDD,small}} \approx 1 + \frac{\alpha a^2 \langle r_m^{\alpha-1} \rangle^2}{\left[\frac{1}{2}\langle r_m^{2\alpha} \rangle + \frac{1}{4}(\alpha^2 + 1)\langle r_s^2 \rangle \langle r_m^{2\alpha-2} \rangle - \alpha a^2 \langle r_m^{\alpha-1} \rangle^2 \right] \frac{f_{BW}}{F_\eta}} \quad (4.50)$$

We can further simplify this expression by noting that r_s is Rayleigh-distributed and therefore $\langle r_s^2 \rangle = 2a^2$. This finally leads to Eq. 4.24.

4.8.2 Approximations for large pure tone signals (CMR)

For the case of pure tone signals, we proceed in the same way as before for the narrow band noise signals. The analytic signal can now be written as

$$\mathcal{A}(t) = r_m e^{i\phi_m} + a e^{i\omega_0 t + i\phi_0}, \quad (4.51)$$

where ϕ_0 is a random initial phase. If the signal amplitude a is large, we approximate the compressed stimulus linearly in r_m , which is now in general small compared to a . For the approximate compressed stimulus, using again Eq. 4.44, one arrives at

$$\begin{aligned} \hat{\gamma}(t) \approx & r_m a^{\alpha-1} \cos \phi_m + a^\alpha \cos(\omega_0 t + \phi_0) + \\ & (\alpha - 1) a^{\alpha-1} r_m \cos(\omega_0 t + \phi_0) \cdot \\ & (\cos \phi_m \cos(\omega_0 t + \phi_0) + \sin \phi_m \sin(\omega_0 t + \phi_0)). \end{aligned} \quad (4.52)$$

For calculating the autocorrelation function, we now also average across the initial phase ϕ_0 (McNamara and Wiesenfeld, 1989) and get

$$C_{\hat{\gamma}\hat{\gamma}}(\tau) \approx \left[\frac{1}{4}(\alpha^2 + 1)\langle r_m^2 \rangle a^{2\alpha-2} \right] \delta(\tau) + \frac{1}{2}a^{2\alpha} \cos \omega_0 \tau. \quad (4.53)$$

Therefore, the onesided power spectral density is

$$S_{\hat{\gamma}\hat{\gamma}}(f)_{\text{CMR,large}} \approx \frac{1}{4}(\alpha^2 + 1)\langle r_m^2 \rangle a^{2\alpha-2} + \frac{1}{2}a^{2\alpha} \delta(f - f_0). \quad (4.54)$$

The numerical signal-to-noise-ratio finally is given by the expression

$$\text{SNR}_{\text{CMR,large}} \approx 1 + \frac{\frac{1}{2}a^{2\alpha}}{\left[\frac{1}{4}(\alpha^2 + 1)\langle r_m^2 \rangle a^{2\alpha-2} \right] \frac{f_{BW}}{F_\eta}} \quad (4.55)$$

4.8.3 Approximations for large narrowband signals and small pure tone signals

The expressions for the power spectral densities in the cases of large narrow band signals and small pure tone signals can be obtained following the same steps as described in Sections 4.8.1 and 4.8.2. They are

$$S_{\hat{\gamma}\hat{\gamma}}(f)_{\text{CDD,large}} \approx \frac{1}{4}(\alpha^2 + 1)\langle r_m^2 \rangle \langle r_s^{2\alpha-2} \rangle + S_{s^{\alpha}s^\alpha}(f) \quad (4.56)$$

and

$$S_{\hat{\gamma}\hat{\gamma}}(f)_{\text{CMR,small}} \approx \frac{1}{2}\langle r_m^{2\alpha} \rangle + \frac{1}{4}(\alpha^2 + 1)a^2 \langle r_m^{2\alpha-2} \rangle - \frac{1}{8}a^2(\alpha + 1)^2 \langle r_m^{\alpha-1} \rangle^2 + \frac{1}{8}a^2(\alpha + 1)^2 \langle r_m^{\alpha-1} \rangle^2 \delta(f - f_0). \quad (4.57)$$

In Eq. 4.56, the term $S_{s^{\alpha}s^\alpha}(f)$ refers to the power spectrum of the compressed signal alone, which can be very well approximated by a box of width f_{BW} and height $2^{\alpha-1}\Gamma(\alpha + 1)a^{2\alpha}/f_{BW}$, using Eq. 4.17 for the standard deviation of compressed Gaussian noise. Remembering that r_s is Rayleigh-distributed, one can furthermore write $\langle r_s^{2\alpha-2} \rangle = 2^{\alpha-1}a^{2\alpha-2}\Gamma(\alpha)$.

CHAPTER 5

SUMMARY AND CONCLUDING REMARKS

It is well known that humans and other animals are better at signal detection in comodulated noise than in unmodulated noise. This is expressed by the investigation of the CMR and CDD effects. For reviews see e.g. Verhey et al. (2003) and Borrill and Moore (2002), respectively. The central question of the present thesis is which mechanisms these effects can be attributed to. By means of neurophysiological data analysis and theoretical considerations, this question has been addressed in the preceding chapters. It turns out that many general aspects of CMR and CDD can be understood by peripheral within-channel processing: a model incorporating frequency filtering, envelope extraction, and compression can reproduce neuronal firing patterns as well as psychophysical signal detection thresholds. Therefore, higher level processing like across-channel comparison seems to be of lesser importance for signal detection in comodulated noise. It could also be shown that just a single compressive processing stage plays an important role for the CDD and CMR effects.

5.1 SUMMARY OF CHAPTERS 2 TO 4

In Chapter 2 neuronal firing rates of avian field L2 neurons in response to CDD stimuli are found to be different for the three correlation conditions AC (all correlated), AU (all uncorrelated), and CU (co-uncorrelated). The CU condition is known to be associated with the lowest signal detection thresholds in psychophysics. One finds that the neural response to the maskers alone is generally higher in the AU case than in the AC and CU cases. With increasing signal level the mean spike discharge rates in the AU and CU cases generally increase monotonically, and both saturate at the same value. By contrast, increasing the signal level in the AC condition leads to a nonmonotonic rate-level function which decreases at first and only starts increasing for signal levels close to the masker level. The differences in the rate-level curves could be attributed to interference due to correlated phases in the CDD stimuli. This result could be reached due to a very simple and therefore largely analytically tractable model of the essential processing stages for reproducing

the firing rates. The model detailed in Chapter 2 incorporates a frequency filtering stage, the extraction of the envelope of the filtered stimulus, and the compression and averaging of the envelope values. In the analytical expressions differences between the correlation conditions could be traced back to interference terms that appear due to phase correlations in the spectra of the stimuli. These interference terms are the reason for the increased masker alone response in the AU condition and for the dip in the AC condition in the model. These two features of the model rate-level curves are present for any compressive exponent $0 < \alpha < 2$, but a realistic compression of $\alpha = 0.35$ was used in the first part of Chapter 2. Employing a simple signal detection scheme (d' based on the assumption of constant standard deviations), the rank order of signal detection thresholds predicted by the model is found to be consistent with psychophysical findings in the literature.

In the second part of Chapter 2 the model processing steps are numerically simulated for the recording sites and the original stimuli used in the experiments in order to predict the neuronal firing rates while fitting as few parameters as possible. It turns out that this numerical model can generally reproduce the neuronal firing rates quite well although there are some differences that may be explained by the fact that the model simulates the signal processing in the auditory system of living animals in a rather abstract way. For example, nonlinearities in the filtering process are not considered, which would lead to suppression, and suppression has been hypothesized to be important for signal detection in comodulated noise (e.g. Moore and Borrill, 2002; Ernst and Verhey, 2006). Despite these drawbacks the dependence of neuronal signal detection thresholds on experimental parameters like signal duration, signal offset, and masker level can largely be reproduced using only the mentioned peripheral processing steps. One possible model prediction for neuronal responses which may be tested with previously published data is that the masker alone response in the case of correlated masker bands should be lower than with uncorrelated masker bands. In Hofer and Klump (2003) the responses of small neural populations of field L2 neurons of European starlings to two narrow noise bands which are either uncorrelated or correlated are reported. For the case in which these two bands are presented at frequencies within the excitatory region of the tuning curve, a difference in response amplitude as predicted by the model can be seen in the data, although it is not statistically significant in all cases.

In Chapter 3 the model introduced in Chapter 2 for a special case of CDD experiments is generalized to be applicable to other CDD experiments and to CMR experiments as well. A survey of the literature shows that CMR and CDD have mostly been modeled separately, possibly due to the prevalent view that CMR has considerable across-channel contributions while CDD can

be understood as a within-channel process. By contrast, the model discussed in the remaining parts of that chapter addresses both CMR and CDD and is based on within-channel processes (although for some cases this within-channel modeling may be interpreted as an effective description of across-channel processes, see Sect. 3.5.1). This psychophysical signal processing scheme is based on neuronal measurements as described in Chapter 2 and is very parsimonious by using only three free parameters. These parameters are the filter bandwidth γ , the compressive exponent α , and the internal noise σ . Despite its simplicity the model can qualitatively explain many aspects of CMR and CDD experiments that are linked to the frequency spacing of noise and signal bands (Sects. 3.4.1 and 3.4.2), the number of flanking bands (Sect. 3.4.3), and the noise band level (Sect. 3.4.4). Therefore, by using the model, one can get an intuitive idea of how and why certain experimental parameters may influence experimental outcomes. This is expressed directly by an analysis of the influence of free parameter changes on model predictions, and indirectly by the findings related to the change of experimental parameters.

Despite the simplicity of the different model variants described in Chapters 2 and 3, it is still possible to further simplify the modeling efforts in order to find the minimal sufficient prerequisites for obtaining the CDD and CMR effects. It has been stated in Sect. 2.4.2 that the general differences between the rate-level curves in the AC, AU, and CU conditions, which are responsible for the CDD effect, are still present without compression (i.e. in the case $\alpha = 1$) and without frequency filtering. That means that the extraction of the envelope alone contributes to the CDD effect and therefore constitutes one minimal sufficient condition for finding improved signal detection in comodulated noise. A second possible sufficient condition for improved signal detection in comodulated noise is discussed in Chapter 4, where the detectability of signals after only one signal processing step, namely compression, is investigated. For a convenient analytical treatment of the problem a generalized form of comodulated noise is defined and analyzed for its properties at the beginning of Chapter 4. Then, as an example for instantaneous compression, envelope compression as used by Bernstein et al. (1999) and van der Heijden (2005) is explained, and the effect of compression on noise intensity is described. By deriving analytical approximations to the signal-to-noise ratio in compressed stimuli, the detectability of signals in compressed comodulated and unmodulated noise is discussed. An increased detectability of signals in comodulated noise compared to uncorrelated noise which is introduced due to the compressive nonlinearity can be explained intuitively when interpreting compression as automatic gain control: the compressive nonlinearity selectively boosts stimulus epochs with relatively low amplitude

while attenuating stimulus epochs with high amplitude. In comodulated noise with small signals, the low intensity epochs are the times at which the signal-to-noise ratio is inherently higher than during times of large intensity. Therefore, the increased signal detectability in comodulated noise can be understood as a dip listening mechanism. In unmodulated noise with small signals, there are no epochs with significantly increased or decreased intensity which means that compression has no positive effect on signal detectability in this case. These findings are generalized to other kinds of compression, such that an important contribution to the CDD and CMR effects can be attributed to the cochlear compressive nonlinearity.

5.2 GENERAL CONCLUSIONS

The main conclusions that can be drawn from the present thesis are summarized in the following five points:

- Enhanced signal detection in comodulated noise as demonstrated by CMR and CDD experiments has at least two important peripheral within-channel contributions: (i) the extraction of the stimulus envelope, which may be localized in the auditory system at the stage where mechanical vibrations of the basilar membrane are transduced to neuronal impulses (spikes), and (ii) the compressive nonlinearity which is a result of active processes in the cochlea mainly generated by the outer hair cells.
- These within-channel contributions to the CDD and CMR effects can be best understood in models which reduce the complex processes in the auditory system to their most essential steps and thus facilitate analytical or semianalytical calculations.
- It is possible to understand many aspects of CDD *and* CMR experiments with a single model which is at the same time very parsimonious (having only three free parameters). Although CDD and CMR have been linked by various authors, such a model has not been developed before.
- A reduction in spike discharge rates with increasing signal level does not necessarily have to be explained by inhibitory or suppressive mechanisms: The dip in the neuronal firing rates obtained for the AC stimuli in CDD experiments can be understood by looking at the statistics of the compressed stimulus envelopes without the need for other mechanisms.

- As detection of signals in comodulated noise does not only pose a challenge for the auditory system but possibly also for other sensory systems or even technical applications, the insights gained in the present work may be generalized and applied to other fields.

5.3 POSSIBILITIES FOR FUTURE RESEARCH

The final point in the conclusions already names one of the main lines of possible future research. Any combination of signal and noise which is transmitted by vibrations (mechanical, electromagnetic or other) can be compared to the situations described in the present thesis. It has been mentioned on several occasions during this thesis that comodulated noise may occur quite frequently in our everyday lives. Therefore, an expansion of the results of the present thesis to other fields (e.g. touch, vision, seismic waves, communication devices, radio astronomy, or even research on gravitational waves) may be interesting, and a thorough analysis of the properties of comodulated noise and its effect on other phenomena like e.g. stochastic resonance is desirable. For example, according to the predictions of the model in this thesis, it might be possible to find the CDD and CMR effects also for tactile vibrations if the envelope of such vibrations is the quantity responsible for their perception.

First results on the comparison between comodulated noise and unmodulated noise for stochastic resonance have been obtained numerically during the course of this thesis. These findings have not been prepared for publication yet, but they may be of importance for signal detection by the auditory system at low sound pressure levels: Zeng et al. (2000) and Ward (2004) discuss the psychoacoustics and psychophysics of stochastic resonance, and Jaramillo and Wiesenfeld (1998) and Henry (1999) investigate physiological evidence for stochastic resonance in the auditory system. The cited references essentially use white noise for their considerations. As comodulated noise is quite ubiquitous in nature, the question arises what influence the presence of this kind of noise has on stochastic resonance. To answer this question, numerical simulations with a standard paradigm of stochastic resonance (*nondynamical* or *threshold* stochastic resonance, see Moss et al., 1994; Gingl et al., 1995) have been performed. It turns out that there is a range of noise levels for which the beneficial effect of noise on the detectability of a signal is larger with comodulated noise than with unmodulated white noise. This is demonstrated in Fig. 5.1, where the signal-to-noise ratio at the output of a thresholding and spiking nonlinearity is compared for unmodulated and comodulated noise. The data in that figure was obtained numerically by

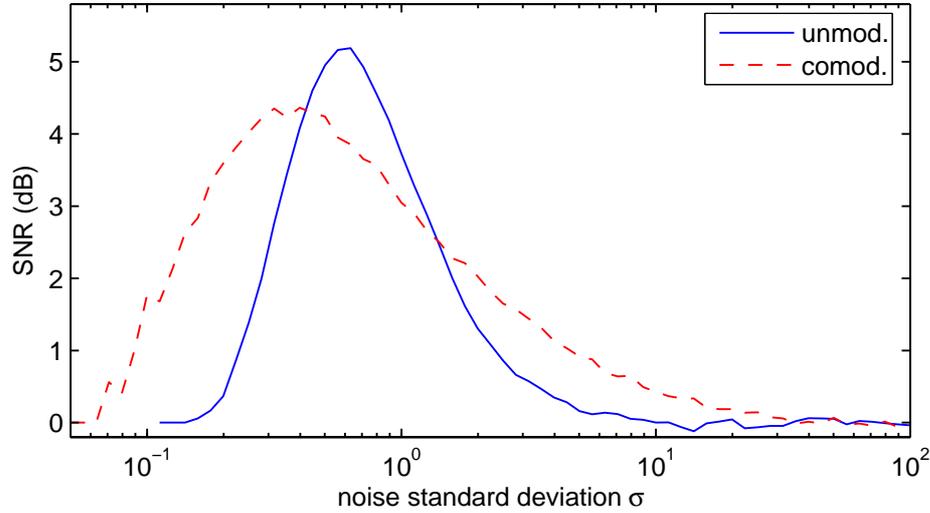


Figure 5.1: Numerical results of simulations for threshold stochastic resonance with unmodulated noise (solid line) and comodulated noise (dashed line). The signal-to-noise ratio as defined in Eq. 4.20 was determined for a spike train generated by eliciting a spike of unit height each time the stimulus (noise of standard deviation σ and sinusoidal signal with amplitude a) crossed a threshold θ from below. Note that due to the definition of the SNR, there is a lower bound for SNR values at 0 dB. (Parameters: $a = 0.5$, $\theta = 1$, $F_p = 50$ Hz, $F_\eta = 5.0$ kHz, $f_0 = 2.0$ kHz, $f_s = 44.1$ kHz.)

the following procedure: A stimulus consisting of comodulated or unmodulated noise as defined in Sect. 4.3 and a sinusoidal signal was converted to a spike train by generating a spike each time the stimulus crossed a constant threshold θ from below. The spikes had unit height and a temporal extent of $1/f_s$, where f_s is the sampling rate used in the simulations. The SNR (signal-to-noise ratio) in this spike train is determined as defined in Eq. 4.20 and is plotted depending on the standard deviation of the noise, keeping a and θ constant. The solid line in Fig. 5.1 shows the well-known characteristics of stochastic resonance: there is a certain noise intensity at which the SNR reaches a maximum. This can be explained by the fact that at this noise intensity, the noise is just enough to drive the stimulus across the threshold each time the signal has a maximum, but the noise intensity is small enough to still keep the spikes locked to the signal's phase. In the case of comodulated noise, the region of increased SNR values extends across a wider range of noise intensities. Although the maximal SNR is smaller than in the case of unmodulated noise, the beneficial effect of the noise is already visible for smaller noise amplitudes, which is due to the fact that comodulated noise has epochs of relatively high noise amplitudes, in which spikes can already be elicited when this would not have been possible with

unmodulated noise. There is also an increased SNR for relatively high noise intensities for comodulated noise compared to unmodulated noise. This can be explained by the fact that even for high noise intensities, comodulated noise has epochs of relatively low noise amplitudes, which now can still elicit spikes that are locked to the signal phase, while this is not possible anymore with unmodulated noise of the same intensity.

These general insights on stochastic resonance with comodulated noise may be more thoroughly investigated in order to formalize the above statements and possibly understand them analytically. The positive influences of comodulated noise on detecting subthreshold signals may also be investigated in psychophysical experiments, where the absolute threshold of hearing might be lowered by the introduction of very low levels of noise (see Zeng et al., 2000; Long et al., 2004, for experimental results with white noise). Stochastic resonance may also be of importance for cochlear implants (e.g. Stocks et al., 2002).

Another promising line of research is a more detailed investigation of the dip found in the firing rates of field L2 neurons in the AC condition of CDD experiments. It would be interesting to know if this dip is also present in the brain response of human listeners and if so, which are the perceptual correlates of that reduced neuronal activity. The first of these two questions is currently being addressed in a research collaboration with André Rupp at the Universitätsklinikum Heidelberg. Using the model described in Chapter 3 with realistic human parameters, a stimulus configuration leading to a strong dip is determined. This AC stimulus configuration and the corresponding CU stimulus (not causing a dip in the model) are then played back to human subjects in an MEG (magnetoencephalography) setup. The measured responses (auditory evoked fields) to both kinds of stimuli may be compared to a reference situation with only the masker bands present, such that the hypothesis can be tested, if the AC stimulus does result in reduced neuronal activity compared to the CU stimulus and the masker alone stimulus. If a reduction in neuronal activity can indeed be observed, psychoacoustical experiments for determining the perceptual consequences of this reduced activity would be highly desirable. One might imagine that the reduction of activity corresponds to a reduced perceived loudness, but preliminary tests do not seem to support this view (unpublished own experiments).

Directly linked to the possible presence of a reduction in neuronal activity for the AC stimulus is yet another line of possible future research: the models used in the present thesis for investigating CDD and CMR stimuli on a neurosensory basis were intentionally kept as simple as possible. It is important to find out if more sophisticated models of neuronal auditory processing such as those used e.g. in Meddis and O'Mard (2005) and Meddis (2006) or

Heinz et al. (2001) can reproduce the model findings of the present thesis (e.g. the dip in the AC condition). Own unpublished data suggests that this is the case, but this question has to be addressed more thoroughly in order to be able to compare the simple model described in Chapter 3 with the more sophisticated models and to identify the main sources of possible differences. This would lead to a better understanding of the origins of reduced activity possibly measured in MEG experiments, and it could contribute to explaining the perceptual correlates of this reduced activity. Apart from that, the influence of suppression and other mechanisms on the neuronal activity in the auditory nerve and higher centers may be investigated using more detailed modeling approaches than those used in the present thesis. For example, suppression may be examined by using the DRNL (dual resonance nonlinear) filter described by Meddis and O'Mard (2001) instead of the filterbank of Hohmann (2002), which was used mostly in the present thesis.

In summary one can say that many questions were addressed and answered during the course of this thesis, but there are still many challenging and exciting questions that wait to be answered. I hope that at least some of the topics mentioned above will be investigated soon.

BIBLIOGRAPHY

- Abraham, D. A., Lyons, A. P., 2004. Simulation of non-Rayleigh reverberation and clutter. *IEEE J. Oceanic Eng.* 29 (2), 347–362.
- Arons, B., 1992. A review of the cocktail party effect. *Journal of the American Voice I/O Society* 12, 35–50.
- Bacon, S. P., Fay, R. R., Popper, A. N. (Eds.), 2003. *Compression: From Cochlea to Cochlear Implants*. Springer, New York.
- Bacon, S. P., Lee, J., 1997. The modulated-unmodulated difference: Effects of signal frequency and masker modulation depth. *J. Acoust. Soc. Am.* 101 (6), 3617–3624.
- Bee, M. A., Buschermöhle, M., Klump, G. M., 2007. Detecting modulated signals in modulated noise: II. Neural thresholds in the songbird forebrain, submitted to *Eur. J. Neurosci.*
- Berg, B. G., 1996. On the relation between comodulation masking release and temporal modulation transfer functions. *J. Acoust. Soc. Am.* 100 (2), 1013–1023.
- Bernstein, L. R., van de Par, S., Trahiotis, C., 1999. The normalized interaural correlation: Accounting for $\text{NoS}\pi$ thresholds obtained with Gaussian and "low-noise" masking noise. *J. Acoust. Soc. Am.* 106 (2), 870–876.
- Borrill, S. J., Moore, B. C. J., 2002. Evidence that comodulation detection differences depend on within-channel mechanisms. *J. Acoust. Soc. Am.* 111 (1), 309–319.
- Bregman, A. S., 1999. *Auditory Scene Analysis: The Perceptual Organization of Sound*, Second MIT Press paperback Edition. MIT Press, Cambridge, MA.
- Bronkhorst, A. W., 2000. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica* 86 (1), 117–128.
- Brumm, H., Slabbekoorn, H., 2005. Acoustic communication in noise. *Adv. Stud. Behav.* 35, 151–209.
- Buchholz, J. M., Mourjopoulos, J., 2004. A computational auditory masking model based on signal-dependent compression. I. Model description and performance analysis. *Acta Acust. United Ac.* 90 (5), 873–886.

- Buschermöhle, M., Feudel, U., Freund, J. A., 2007a. Enhanced signal detectability in comodulated noise introduced by compression, submitted to J. Acoust. Soc. Am.
- Buschermöhle, M., Feudel, U., Klump, G. M., Bee, M. A., Freund, J. A., 2006. Signal detection enhanced by comodulated noise. *Fluct. Noise Lett.* 6 (4), L339–L347.
- Buschermöhle, M., Verhey, J. L., Feudel, U., Freund, J. A., 2007b. A unifying model approach to comodulation detection difference (CDD) and comodulation masking release (CMR), submitted to Biol. Cybern.
- Buus, S., 1985. Release from masking caused by envelope fluctuations. *J. Acoust. Soc. Am.* 78 (6), 1958–1965.
- Buus, S., Klump, G. M., Gleich, O., Langemann, U., 1995. An excitation-pattern model for the starling (*Sturnus vulgaris*). *J. Acoust. Soc. Am.* 98 (1), 112–124.
- Carlyon, R. P., Jaysurya Datta, A., 1997. Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system. *J. Acoust. Soc. Am.* 101 (6), 3636–3647.
- Cohen, M. F., 1991. Comodulation masking release over a three octave range. *J. Acoust. Soc. Am.* 90 (3), 1381–1384.
- Cohen, M. F., Schubert, E. D., 1987. The effect of cross-spectrum correlation on the detectability of a noise band. *J. Acoust. Soc. Am.* 81, 721–723.
- Dallos, P., 1992. The active cochlea. *J. Neurosci.* 12 (12), 4575–4585.
- Dau, T., Kollmeier, B., Kohlrausch, A., 1997. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.* 102 (5), 2892–2905.
- Embleton, T. F. W., 1996. Tutorial on sound propagation outdoors. *J. Acoust. Soc. Am.* 100 (1), 31–48.
- Ernst, S., Verhey, J. L., 2005. Comodulation masking release over a three octave range. *Acta Acust. United Ac.* 91, 998–1006.
- Ernst, S. M. A., Verhey, J. L., 2006. Role of suppression and retro-cochlear processes in comodulation masking release. *J. Acoust. Soc. Am.* 120 (6), 3843–3852.
- Fantini, D. A., Moore, B. C. J., 1994. Profile analysis and comodulation detection differences using narrow bands of noise and their relation to comodulation masking release. *J. Acoust. Soc. Am.* 95 (4), 2180–2191.

- Fishbach, A., May, B. J., 2003. A neural edge-detection model for enhanced auditory sensitivity in modulated noise. In: Becker, S., Thrun, S., Obermayer, K. (Eds.), *Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA, pp. 285–292.
- Gabor, D., 1946. Theory of communication. *J. IEE (London)* 93, 429–457.
- Gardiner, C. W., 2004. *Handbook of Stochastic Methods*. Springer.
- Gingl, Z., Kiss, L. B., Moss, F., 1995. Non-dynamical stochastic resonance: Theory and experiments with white and arbitrarily coloured noise. *Europhys. Lett.* 29 (3), 191–196.
- Glasberg, B. R., Moore, B. C. J., 1990. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138.
- Gleich, O., 1994. Excitation patterns in the starling cochlea: A population study of primary auditory afferents. *J. Acoust. Soc. Am.* 95 (1), 401–409.
- Green, D. M., Swets, J. A., 1966. *Signal detection theory and psychophysics*. Wiley, New York.
- Griffiths, T. D., Warren, J. D., 2004. What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892.
- Grose, J. H., Hall, J. W., 1996. Cochlear hearing loss and the processing of modulation: Effects of temporal asynchrony. *J. Acoust. Soc. Am.* 100 (1), 519–527.
- Hall, J. W., Buss, E., Grose, J. H., 2006. Comodulation detection differences for fixed-frequency and roved-frequency maskers. *J. Acoust. Soc. Am.* 119 (2), 1021–1028.
- Hall, J. W., Davis, A. C., Haggard, M. P., Pillsbury, H. C., 1988. Spectro-temporal analysis in normal-hearing and cochlear-impaired listeners. *J. Acoust. Soc. Am.* 84 (4), 1325–1331.
- Hall, J. W., Grose, J. H., 1994. Signal detection in complex comodulated backgrounds by normal hearing and cochlear-impaired listeners. *J. Acoust. Soc. Am.* 95 (1), 435–443.
- Hall, J. W., Haggard, M. P., Fernandes, M. A., 1984. Detection in noise by spectro-temporal pattern analysis. *J. Acoust. Soc. Am.* 76 (1), 50–56.
- Harte, J. M., Elliott, S. J., Rice, H. J., 2005. A comparison of various nonlinear models of cochlear compression. *J. Acoust. Soc. Am.* 117 (6), 3777–3786.
- Hartmann, W. M., 1998. *Signals, Sound, and Sensation*. Springer, New York.

- Heinz, M. G., Colburn, H. S., Carney, L. H., 2001. Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Comput.* 13 (10), 2273–2316.
- Henry, K. R., 1999. Noise improves transfer of near-threshold, phase-locked activity of the cochlear nerve: Evidence for stochastic resonance? *J. Com. Physiol. A* 184, 577–584.
- Hofer, S. B., Klump, G. M., 2003. Within- and across-channel processing in auditory masking: A physiological study in the songbird forebrain. *J. Neurosci.* 23 (13), 5732–5739.
- Hohmann, V., 2002. Frequency analysis and synthesis using a gammatone filterbank. *Acta Acust. United Ac.* 88, 433–442.
- Jaramillo, F., Wiesenfeld, K., 1998. Mechano-electrical transduction assisted by Brownian motion: A role for noise in the auditory system. *Nature Neurosci.* 1 (5), 384–388.
- Jensen, K. K., 2007. Comodulation detection differences in the hooded crow (*Corvus corone cornix*), with direct comparison to human subjects. *J. Acoust. Soc. Am.* 121 (3), 1783–1789.
- Joris, P. X., Schreiner, C. E., Rees, A., 2004. Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84 (2), 541–577.
- Klump, G. M., 1996. In: Kroodsma, D. E., Miller, E. H. (Eds.), *Ecology and Evolution of Acoustic Communication in Birds*. Cornell University Press, Ithaca.
- Klump, G. M., Nieder, A., 2001. Release from masking in fluctuating background noise in a songbird's auditory forebrain. *NeuroReport* 12 (9), 1825–1829.
- Köppel, C., Yates, G., 1999. Coding of sound pressure level in the barn owl's auditory nerve. *J. Neurosci.* 19 (21), 9674–9686.
- Langemann, U., Bee, M. A., Klump, G. M., 2005. Comodulation effects in a songbird: Comodulation detection difference. In: *Assoc. Res. Otolaryngol. Abs.* p. 253.
- Langemann, U., Klump, G. M., 2001. Signal detection in amplitude-modulated maskers. I. Behavioural auditory thresholds in a songbird. *Eur. J. of Neurosci.* 13, 1025–1032.
- Langemann, U., Klump, G. M., 2007. Detecting modulated signals in modulated noise: I. Behavioural auditory thresholds in a songbird, submitted to *Eur. J. Neurosci.*

- Las, L., Stern, E. A., Nelken, I., 2005. Representation of tone in fluctuating maskers in the ascending auditory system. *J. Neurosci.* 25 (6), 1503–1513.
- Lawson, J. L., Uhlenbeck, G. E., 1950. *Threshold Signals*. McGraw-Hill, New York.
- Long, Z.-C., Shao, F., Zhang, Y.-P., Qin, Y.-G., 2004. Noise-enhanced hearing sensitivity. *Phys. Lett. A* 323, 434–438.
- Lopez-Poveda, E. A., Meddis, R., 2001. A human nonlinear cochlear filterbank. *J. Acoust. Soc. Am.* 110 (6), 3107–3118.
- Manley, G. A., Popper, A. N., Fay, R. R. (Eds.), 2004. *Evolution of the Vertebrate Auditory System*. Springer-Verlag, New York.
- McFadden, D., 1987. Comodulation detection differences using noise-band signals. *J. Acoust. Soc. Am.* 81 (5), 1519–1527.
- McNamara, B., Wiesenfeld, K., 1989. Theory of stochastic resonance. *Phys. Rev. A* 39, 4854–4869.
- Meddis, R., 2006. Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. *J. Acoust. Soc. Am.* 119 (1), 406–417.
- Meddis, R., Delahaye, R., O'Mard, L., Sumner, C., Fantini, D. A., Winter, I., Pressnitzer, D., 2002. A model of signal processing in the cochlear nucleus: Comodulation masking release. *Acta Acust. United Ac.* 88 (3), 387–398.
- Meddis, R., O'Mard, L. P., 2001. A computational algorithm for computing nonlinear auditory frequency selectivity. *J. Acoust. Soc. Am.* 109 (6), 2852–2861.
- Meddis, R., O'Mard, L. P., 2005. A computer model of the auditory-nerve response to forward-masking stimuli. *J. Acoust. Soc. Am.* 117 (6), 3787–3798.
- Moore, B. C. J., 2003. *An Introduction to the Psychology of Hearing, Fifth Edition*. Academic Press, London.
- Moore, B. C. J., Borrill, S. J., 2002. Tests of a within-channel account of comodulation detection differences. *J. Acoust. Soc. Am.* 112 (5), 2099–2109.
- Moore, B. C. J., Shailer, M. J., 1991. Comodulation masking release as a function of level. *J. Acoust. Soc. Am.* 90 (2), 829–835.
- Moore, B. C. J., Shailer, M. J., Hall, J. W., Schooneveldt, G. P., 1993. Comodulation masking release in subjects with unilateral and bilateral hearing impairment. *J. Acoust. Soc. Am.* 93 (1), 435–451.
- Moss, F., Pierson, D., O'Gorman, D., 1994. Stochastic resonance: Tutorial and update. *Int. Bifurc. Chaos* 4 (6), 1383–1397.

- Nelken, I., Rotman, Y., Yosef, O. B., 1999. Response of auditory-cortex neurons to structural features of natural sounds. *Nature* 397, 154–157.
- Neuert, V., Verhey, J. L., 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., Klump, G. M., 1999. Adjustable frequency selectivity of auditory fore-brain neurons recorded in a freely moving songbird via radiotelemetry. *Hearing Research* 127, 41–54.
- O’Loughlin, B. J., Moore, B. C. J., 1981. Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking. *J. Acoust. Soc. Am.* 69 (4), 1119–1125.
- Oxenham, A. J., Bacon, S. P., 2003. Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear and Hearing* 24 (5), 352–366.
- Papoulis, A., Unnikrishna Pillai, S., 2002. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York.
- Patterson, R. D., Nimmo-Smith, I., 1980. Off-frequency listening and auditory-filter asymmetry. *J. Acoust. Soc. Am.* 67 (1), 229–245.
- Pickles, J. O., 1988. *An Introduction to the Physiology of Hearing*, 2nd Edition. Academic Press.
- Plack, C., 2004. Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss. *J. Acoust. Soc. Am.* 115 (4), 1684–1695.
- Plack, C. J., Oxenham, A. J., Drga, V., 2002. Linear and nonlinear processes in temporal masking. *Acta Acust. United Ac.* 88, 348–358.
- Pressnitzer, D., Meddis, R., Delahaye, R., Winter, I. M., 2001. Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. *The Journal of Neuroscience* 21 (16), 6377–6386.
- Priestley, M. B., 1983. *Spectral Analysis and Time Series*. Academic Press.
- Rhode, W. S., Recio, A., 2000. Study of mechanical motions in the basal region of the chinchilla cochlea. *J. Acoust. Soc. Am.* 107 (6), 3317–3332.
- Richards, D. G., Wiley, R. H., 1980. Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication. *Am. Nat.* 115 (3), 381–399.

- Richards, V. M., 1987. Monaural envelope correlation perception. *J. Acoust. Soc. Am.* 82 (5), 1621–1630.
- Rieke, F., Bodnar, D. A., Bialek, W., 1995. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B* 262 (1365), 259–265.
- Robles, L., Ruggero, M. A., 2001. Mechanics of the mammalian cochlea. *Physiol. Rev.* 81 (3), 1305–1352.
- Rosengard, P. S., Oxenham, A. J., Braida, L. D., 2005. Comparing different estimates of cochlear compression in listeners with normal and impaired hearing. *J. Acoust. Soc. Am.* 117 (5), 3028–3041.
- Schooneveldt, G. P., Moore, B. C. J., 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., Moore, B. C. J., 1989. Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands. *J. Acoust. Soc. Am.* 85 (1), 262–272.
- Schreiner, C. E., Urbas, J. V., 1988. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear. Res.* 32 (1), 49–63.
- Singh, N. C., Theunissen, F. E., 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* 114 (6), 3394–3411.
- Smith, E. C., Lewicki, M. S., 2006. Efficient auditory coding. *Nature* 439, 978–982.
- Stocks, N. G., Allingham, D., Morse, R. P., 2002. The application of suprathreshold stochastic resonance to cochlear implant coding. *Fluct. Noise Lett.* 2 (3), 169–181.
- Sumner, C. J., Lopez-Poveda, E. A., O’Mard, L. P., Meddis, R., 2002. A revised model of the inner-hair cell and auditory-nerve complex. *J. Acoust. Soc. Am.* 111 (5), 2178–2188.
- Tatarski, V. I., 1961. *Wave propagation in a turbulent medium*. McGraw-Hill, New York.
- van de Par, S., Kohlrausch, A., 1998a. Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research. *J. Acoust. Soc. Am.* 103 (6), 3605–3620.

- van de Par, S., Kohlrausch, A., 1998b. Diotic and dichotic detection using multiplied noise maskers. *J. Acoust. Soc. Am.* 103 (4), 2100–2110.
- van der Heijden, M., 2005. Cochlear gain control. *J. Acoust. Soc. Am.* 117 (3), 1223–1233.
- Verhey, J. L., Dau, T., 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., Winter, I. M., 2003. The psychophysics and physiology of comodulation masking release. *Exp. Brain Res.* 153, 405–417.
- Verhey, J. L., Rennies, J., Ernst, S. M. A., 2007. Influence of envelope distributions on signal detection, accepted for publication in *Acta Acust. United Ac.*
- Ward, L. M., 2004. Psychophysics of stochastic resonance. *Fluct. Noise Lett.* 4 (1), L11–L21.
- Weisstein, E. W., 2002. *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall/CRC.
- Wiley, R. H., Richards, D. G., 1982. *Acoustic Communication in Birds*. Academic Press, New York, Ch. Adaptations for Acoustic Communication in Birds: Sound Transmission and Signal Detection, pp. 131–181.
- Wright, B. A., 1990. Comodulation detection differences with multiple signal bands. *J. Acoust. Soc. Am.* 87 (1), 292–303.
- Xu, Z. W., Wu, J., Wu, Z. S., 2004. A survey of ionospheric effects on space-based radar. *Wave Random Media* 14, 189–273.
- Zeng, F.-G., Fu, Q.-J., Morse, R., 2000. Human hearing enhanced by noise. *Brain Res.* 869, 251–255.

ACKNOWLEDGMENTS

After nearly three years of work at the International Graduate School for Neurosensory Science, Systems, and Applications (DFG Graduiertenkolleg 591, InterGK) as well as at the Institute for Chemistry and Biology of the Marine Environment (ICBM), I can say that it was a pleasure to collaborate and interact with colleagues (graduates and senior scientists) from diverse fields of research. It will be hard to mention everyone by name who contributed in one way or the other to the successful completion of the present work. While Ulrike Feudel as my thesis advisor provided me with a lot of support for the writing of manuscripts and the work on the theoretical concepts, I also enjoyed the extremely helpful discussions with Jan Freund about stochastic processes, signal processing, and neuronal modeling. Of course I would not have been able to begin my theoretical work without the kind supply of experimental data by Mark Bee and Georg Klump who both also supported me by way of discussion and by explaining many of the relevant physiological details to me. Already early on during my time at the InterGK, discussions with Jesko Verhey guided me on the right tracks for my investigations, and still these discussions provide me with new insights and exciting ideas. A further partner for relevant conversations was Ulrike Lange-
mann, who taught me a lot about animal psychophysics. André Rupp at the Universitätsklinikum Heidelberg was a further partner for discussions and is involved in ongoing collaborations. And I do not want to miss the chance of thanking Birger Kollmeier for being one of the driving forces of hearing research in general and of the graduate school in particular and therefore for providing an extremely stimulating scientific environment.

Certainly, the present thesis would have been completely different (i.e. worse) without the innumerable discussions and meetings with other graduate students directly or indirectly involved in the InterGK. I would like to name a few of the most important by name, although everyone has had his or her share in improving my work or my working conditions, respectively: I learned many techniques and background information regarding psychophysical experiments from Stephan Heise and Marc Nitschmann. The importance of compression was pointed out to me among others by Suzan Emiroglu. My technical questions could nearly always be answered by Stefan Strahl and Rainer Beutelmann. My knowledge of the physiology of the auditory sys-

tem was constantly increased by Melanie Zokoll, Karin Klink, Julia Maier, and Astrid Klinge. I discussed the possibilities of comodulation experiments in the tactile domain with Melina Brell. Stephan Ernst shared a lot of his knowledge on the modeling and psychophysics of CDD and CMR with me. And Helge Lüddemann told me about the implementation of compression which I used in Chapter 4. Of course I also really enjoyed the private and professional meetings with all those graduates whom I have not mentioned here.

I also wish to thank all the members of the complex systems research group at the ICBM who were always at hand when I needed help. I am especially grateful to Jöran März who helped me a lot with typesetting this work in \LaTeX , to Klemens Buhmann who assisted me in the case of computer problems, and to everyone else for the nice coffee breaks.

Last but not least I wish to express my deepest gratitude to all my friends and family who accompanied me along my way. My parents Agnes and Franz, my brothers Andreas and Frank as well as my parents in law, and all my friends have provided help and comfort whenever necessary. And of course, finally, I would not be what I am today without my loving wife Katrin who clearly knows best how to keep me motivated with my work and how to distract me from too much science if necessary. I love you very much!

LEBENS LAUF

Michael Buschermöhle
geboren am 09. Juni 1977 in
Wildeshausen
Staatsangehörigkeit: deutsch



- seit Juni 2004: Promotionsstudium der theoretischen Physik an der Carl von Ossietzky Universität Oldenburg; Stipendiat im Internationalen Graduiertenkolleg für Neurosensorik
- August 2003 bis Mai 2004: Master of Science am Rensselaer Polytechnic Institute (Troy, New York)
Schwerpunkt: Astrophysik/Ursprünge des Lebens
- Juli 2003: Diplom-Physiker (Max-Planck-Institut für Strömungsforschung Göttingen, Georg-August-Universität Göttingen)
Titel der Diplomarbeit: *Latenzvariabilität synchroner Spikeaktivität in kortikalen Netzwerken*
- Oktober 1998 bis Juli 2003: Studium der Physik an der Georg-August-Universität Göttingen
- Juli 1997 bis September 1998: Grundwehrdienst
- Juni 1997: Abitur am Clemens-August-Gymnasium Cloppenburg

ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst habe und nur die angegebenen Hilfsmittel verwendet habe. Teile der Dissertation wurden bereits veröffentlicht bzw. sind zur Veröffentlichung eingereicht, wie an den entsprechenden Stellen angegeben. Die Dissertation hat weder in Teilen noch in ihrer Gesamtheit einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen.

Oldenburg, den 07. Mai 2007.

.....
Michael Buschermöhle