

Modeling auditory processing of amplitude modulation

Torsten Dau



**Bibliotheks- und Informationssystem der Universität Oldenburg
1996**

Verlag/Druck/
Vertrieb

Bibliotheks- und Informationssystem
der Carl von Ossietzky Universität Oldenburg
(BIS) - Verlag -
Postfach 2541, 26015 Oldenburg
Tel.: 0441/798 2261, Telefax: 0441/798 4040
e-mail: sip@bis1.uni-oldenburg.de

ISBN 3-8142-0570-7

Modeling auditory processing of amplitude modulation

Torsten Dau

Preface

A detailed knowledge of the processes involved in hearing is an essential prerequisite for numerous medical and technical applications, such as, e.g., diagnosis and treatment of hearing disorders, construction and fitting of digital hearing aids, public address systems in theaters and other auditoria, and speech processing in telecommunication and man-machine interaction. Although much is known about the physiology and psychology of hearing as well as the “effective” signal processing in the auditory system, still many unsolved problems remain and even more fascinating properties of the human ear still have to be characterized by the scientist. This is one of the primary goals of the interdisciplinary graduate college “Psychoacoustics” at the University of Oldenburg where physicists, psychologists, computer scientists, and physicians (specialized in audiology) pursue an interdisciplinary approach towards a better understanding of hearing and its various applications. Within this graduate college, approximately 25 Ph.D. students perform their respective Ph.D. work and training program in an interdisciplinary context. The current issue is based on the doctoral dissertation by Torsten Dau and is one of the most outstanding “outputs” of this graduate college.

Torsten Dau’s work is focussed on the quantitative modeling of the auditory system’s performance in psychoacoustical experiments. Rather than trying to model each physiological detail of auditory processing, his approach is to focus on the “effective” signal processing in the auditory system which uses as little physiological assumptions and physical parameters as necessary, but tries to predict as many psychoacoustical aspects and effects as possible. While his previous work has focussed on temporal effects of auditory processing, Torsten Dau’s dissertation focuses on the perception and processing of amplitude modulations. This topic is of particular importance, because most of the natural signals (including speech) are characterized by amplitude modulations and, in addition, physiological data provide evidence of specialized amplitude modulation processing systems in the brain. Thus, an adequate modeling of modulation perception should be a key to the quantitative understanding of the functioning of our ear. The current work now presents a new quantitative signal processing model and validates this model by using “critical” experiments both from the literature and by using data from own experiments.

The main chapters of the current work (chapters 2-4) are self-consistent papers that have already been submitted in a modified version to scientific journals. The first of these main parts (chapter 2) develops the structure of the processing model by developing a kind of “artificial” listener, i.e., a computer model which is fed by the same signals as in the psychoacoustical experiments performed with human listeners and is constructed to predict the responses on a trial-by-trial basis. The specialty of this model is the modulation filterbank which forms an essential improvement over previous versions of the model. The current modeling approach reflects the close cooperation between the research groups at the “Drittes Physikalisches Institut” in Göttingen, the IPO in Eindhoven, and the University of Oldenburg, and is based on many years of experience in psychoacoustic research. With this modulation filterbank, several effects of modulation detection and modulation masking can be explained in a very exact and intriguing way. In addition, analytical calculations are presented that deal with the modulation spectra of bandpass-filtered signals. Also, an extensive comparison is made between own measurements and model predictions and results from the literature. Thus, a large body of data and several compelling arguments are collected that favour the model structure developed here.

Chapter 3 extends the model which was originally designed to deal with narrow-band signals to the important case of broad-band signals and the case of considering a larger temporal range. The intriguing “trick” used by Torsten Dau is to simultaneously evaluate several auditory channels with a combined “optimum” detector so that an equivalence exists between the evaluation of several narrow-band signals and a single broad-band signal. Since previous models of modulation processing from the literature assume such a broad-band analysis, this approach bridges the gap between these previous models and the model developed here. A similar principle is used for the temporal domain where the temporal extension of the signal yields a better detectability of amplitude modulations. This increase in detectability can be described in an intriguing way by appropriate choice of the optimum detector. This concept thus yields a mathematical formulation of the “multiple-look strategy” often referred to in the literature. As in the previous chapter, Torsten Dau can predict both the own experimental data and the data from the literature.

The fourth chapter finally deals with the special case of amplitude modulation of sinusoidal carriers at very high frequencies where the coding of information in the central nervous system does not allow for a unique temporal representation of acoustical signals. Because of this effect, previous studies from the literature could not describe the results of modulation perception experiments in a satisfactory way. Torsten Dau can now show in a very impressive way that his model structure is also capable of explaining these experimental data. Although the coincidence between his predictions and the data is not as “perfect” as in the previous chapters, the possible causes for these discrepancies are explained in detail.

Taken together, the current work can be considered an important milestone in the quantitative description of the effective signal processing in the auditory system. Based on this modeling approach introduced here, the science of psychoacoustics can be put on a quantitative, numerical foundation. Thus, it might eventually be possible to distinguish between “processing” factors and “psychological” factors contributing to the hearing process. These “processing” factors can be incorporated in a “computer ear” which might be the basis for future applications such as digital hearing aids, speech coders, and speech recognition systems. Thus, the current work seems to be both of interest to fundamental scientists (who are seeking to understand the functioning of the highly nonlinear and complex human auditory system) and to applied scientists (who seek to use auditory principles for the improvement of technical systems in hearing and speech technology).

I hope that the reader will enjoy reading this work in a similar way as I enjoyed working with Torsten on his dissertation and that the reader might get some impression of the truly interdisciplinary spirit of the graduate college in Oldenburg.

Oldenburg, summer 1996

Birger Kollmeier

Modeling auditory processing of amplitude modulation

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Torsten Dau
geb. am 28.03.1965
in Hannover

Erstreferent: Prof. Dr. Dr. Birger Kollmeier
1. Korreferent: Prof. Dr. Volker Mellert
2. Korreferent: Dr. Armin Kohlrausch
Tag der Disputation: 2. Februar 1996

Abstract

In this thesis a new modeling approach is developed which is able to predict human performance in a variety of experimental conditions related to modulation detection and modulation masking. Envelope fluctuations are analyzed with a modulation filterbank. The parameters of the filterbank were adjusted to allow the model to account for modulation detection and modulation masking data with narrowband carriers at a high center frequency. In the range 0-10 Hz, the modulation filters have a constant bandwidth of 5 Hz. Between 10 and 1000 Hz a logarithmic scaling with a constant Q-value of 2 is assumed. This leads to the following predictions: For conditions in which the modulation frequency (f_{mod}) is smaller than half the bandwidth of the carrier (Δf), the model predicts an increase in modulation thresholds with increasing modulation frequency. This prediction agrees with the lowpass characteristic in the temporal modulation transfer function (TMTF) in the literature. Within the model this lowpass characteristic is caused by the logarithmic scaling of the modulation filter bandwidth. In conditions with $f_{mod} > \frac{\Delta f}{2}$, the model can account for the highpass characteristic in the threshold function, reflecting the auditory system's frequency selectivity for modulation.

In modulation detection conditions with carrier bandwidths larger than a critical band, the modulation analysis is performed in parallel within each excited peripheral channel. In the detection stage of the model, the outputs of all modulation filters from all excited peripheral channels are combined linearly and with optimal weights. The model accounts for the findings that, (i), the "time constants" associated with the temporal modulation transfer functions (TMTFs) for bandlimited noise carriers do not vary with carrier center frequency and that, (ii), the time constants associated with the TMTF's decrease monotonically with increasing carrier bandwidth. The model also accounts for data of modulation masking with broadband noise carriers. The predicted masking pattern produced by a narrowband noise along the modulation frequency scale is in very good agreement with results from the literature.

To integrate information across time, a "multiple-look" strategy is realized within the detection stage. This strategy allows the model to account for long time constants derived from the data on modulation integration without introducing true long-term integration. Instead, the long "effective" time constants result from the combination of information from different "looks" via multiple sampling and probability summation.

In modulation detection experiments with deterministic carriers (such as sinusoids), the limiting factor for detecting modulation within the model is the *internal* noise that is added as independent noise to the output of all modulation filters in all peripheral filters. In addition, the shape of the peripheral filters plays a major role in stimulus conditions where the detection is based on the "audibility" of the spectral sidebands of the modulation. The model can account for the observed flat modulation detection thresholds up to a modulation rate of about 100 Hz and also for the frequency-dependent roll-off in the threshold function observed in the data for a set of carrier frequencies in the range from 2–9 kHz.

The model might also be used in applications such as psychoacoustical experiments with hearing-impaired listeners, speech intelligibility and speech quality predictions.

Contents

1	General Introduction	3
2	Modulation detection and masking with narrowband carriers	7
2.1	Introduction	9
2.2	Description of the model	13
2.2.1	Original model of the “effective” signal processing	13
2.2.2	Extension of the model for describing modulation perception	15
2.3	Envelope statistics and envelope spectra of Gaussian noises	20
2.4	Method	24
2.4.1	Procedure and Subjects	24
2.4.2	Apparatus and stimuli	24
2.5	Results	26
2.5.1	Measurements and simulations of modulation detection and modulation masking	26
2.5.2	Link between modulation detection and intensity discrimination	35
2.6	Predictions of Viemeister’s model for modulation detection	40
2.7	Discussion	44
2.8	Conclusions	47
3	Spectral and temporal integration in modulation detection	48
3.1	Introduction	50
3.2	Method	52
3.2.1	Procedure and Subjects	52
3.2.2	Apparatus and stimuli	52
3.3	Multi-channel model	53
3.4	Results from measurements and simulations	55
3.4.1	Modulation analysis within and beyond one critical band	55
3.4.2	Effects of bandwidth and frequency region	56
3.4.3	Further experiments and analytical considerations	61
3.4.4	Predictions for modulation masking using broadband noise carriers	67
3.4.5	Temporal integration in modulation detection	71

3.5	Discussion	74
3.5.1	Spectral integration	74
3.5.2	Temporal integration	75
3.5.3	Future extensions of the model	78
3.6	Conclusions	80
4	Amplitude modulation detection with sinusoidal carriers	81
4.1	Introduction	83
4.2	Method	85
4.2.1	Procedure and Subjects	85
4.2.2	Apparatus and stimuli	85
4.3	Experimental results and model predictions	86
4.3.1	Amplitude modulation detection thresholds for a carrier frequency of 5 kHz	86
4.3.2	Comparison of sideband detection and amplitude modulation detection data	87
4.3.3	Simulations on the basis of the modulation filterbank model	89
4.3.4	TMTFs for different carrier frequencies	92
4.4	Discussion	95
4.5	Conclusions	100
5	Summary and conclusion	101
A	Contributions from Signal detection theory (SDT)	104
A.1	Formal discussion of the decision problem	104
A.2	The decision problem in an <i>m</i> IFC task	106
A.3	Gaussian assumption and the probability of correct decisions . . .	107
B	Transformation of the nonlinear adaptation circuits	110
	References	112
	Danksagung	121
	Lebenslauf	122

Chapter 1

General Introduction

The auditory system provides us with access to a wealth of acoustic information, performing a complex transform of the sound energy incident at our ears into percepts which enable us to orient ourselves and other objects within our surroundings. A major aim of psychoacoustic research is to establish functional relationships between the basic physical attributes of sound, such as intensity, frequency and changes in these characteristics over time, and their associated percepts. Quantitative studies, using tasks designed to measure behavioral thresholds for the detection and discrimination of various stimuli, assist us in this aim. This study deals particularly with the dimension of *time* in auditory processing.

With most sounds in our environment, such as speech and music, information is contained to a large extent in the changes of sound parameters with time, rather than in the stationary sound segments. We might therefore expect that the auditory system is able to follow temporal variations to a high degree of accuracy. Methods of quantifying the temporal resolution of the auditory system include measuring the ability of listeners to detect a brief temporal gap between two stimuli, or to detect a sound that is modulated in some way. Compared with other sensory systems, the auditory system is “fast”, in that we are able to hear temporal changes in the range of a few milliseconds and can hear the perceptual “roughness” produced by periodically interrupting a broadband noise at a rate of up to several kilohertz. This ability is several orders of magnitude faster than in vision, where modulations in intensity greater than 60 Hz go unnoticed.

When discussing temporal variations, it is necessary to distinguish between the *fine structure* of the sound, i.e., the variations in instantaneous pressure, and the *envelope* of the sound, i.e., the slower, overall changes in the amplitude. In psychoacoustics, temporal resolution normally refers to the latter (e.g. Viemeister and Plack, 1993).

It is commonly assumed that two general sources of temporal resolution limitation in the auditory system can be distinguished: those of “peripheral” and those of “central” origin. The term peripheral is associated with the first stages of auditory processing, up to and including the processing in the auditory nerve.

These stages include the filtering of the basilar membrane which necessarily influences temporal resolution: Temporal fluctuations which occur with a higher rate than the bandwidth of the auditory filter will be attenuated by the transfer function of the filter. Due to the variation in auditory filter bandwidth with frequency, this limitation to temporal resolution should be frequency dependent. It will affect low-frequency sounds much more strongly than high-frequency sounds. Second, the properties of hair cells, synapses, and the refractory period of neurons limit the maximal discharge rate that can be achieved in the auditory nerve. This limits the rate of envelope fluctuations that can be encoded. This influence will be *similar* at all stimulus frequencies.

Central limitations of the temporal resolution may result from the processing of information at higher stages in the auditory pathway. When measuring thresholds for detecting fluctuations in the amplitude of a sound as a function of the rate of fluctuation, it is observed that thresholds progressively increase with increasing modulation rate (e.g., Viemeister, 1979). The system seems to become less sensitive to amplitude modulation as the rate of modulation increases. Since the response of the peripheral stages at high frequencies should be too fast to be the limiting factor, this has led to the idea that there is a process at a higher level which is “sluggish” in some way (e.g., Moore and Glasberg, 1986). Models of temporal resolution are especially concerned with this process.

There is a very popular type of model described in the literature, which has been developed for describing temporal resolution (e.g., Viemeister, 1979). This model consists of the following stages: (i) bandpass filtering, (ii) a rectifying non-linearity, (iii) a lowpass filter and (iv) a decision mechanism (for a review, see Viemeister and Plack, 1993). The bandpass filtering corresponds to peripheral filtering. The nonlinearity (e.g., half-wave rectification) introduces low-frequency components corresponding to the envelope of the signal. The next stage of low-pass filtering (or integration) is intended to simulate the temporal resolution limit by attenuating rapid changes in the envelope of the signal. The decision mechanism is intended to simulate how the subject uses the output of the integrator to make a discrimination in a specific task. A variety of decision algorithms has been used for this model: the signal-to-noise ratio at a particular time in the stimulus (Moore et al., 1988), the overall variance of the output of the integrator (Viemeister, 1979), or the ratio between the maximum and minimum values of the output (Forrest and Green, 1987).

The present study describes a model which differs considerably from the above modeling approaches. The work builds up on many years of modeling work started about 10 years ago in the psychoacoustic research group at the University of Göttingen. The model includes as an important part a nonlinear adaptation stage which simulates adaptive properties of the periphery and enables the model to account for data of forward masking (Püschel, 1988). Another further stage, which also differs considerably from the models described above, is the decision mechanism. It is implemented as an “optimal detector” which performs some kind

of pattern recognition of the whole temporal course of the internal representation of the stimuli (Dau, 1992; Dau et al., 1995a). This behavior is in contrast to the detection mechanisms in the Viemeister model, which are based on a particular point in time, or on a simple averaging process across time.

This thesis is concerned with the extension of the “effective signal processing” of the auditory system to conditions of modulation detection and modulation masking. As a substantially new part of signal processing, a modulation filterbank is introduced to analyze the envelope fluctuations of the stimuli in each peripheral auditory filter. The inclusion of a modulation filterbank, which presumably represents processing at stages higher than the auditory nerve, is motivated by results from several studies on modulation masking (e.g., Kay and Green, 1973, 1974; Martens, 1982; Bacon and Grantham, 1989; Houtgast, 1989) and recent data and model predictions from Fassel and Püschel (1993), Münkner (1993a+b) and Fassel (1994). The authors suggested modulation channels to account for effects of frequency selectivity in the modulation frequency domain. Apart from the study of Fassel (1994) who investigates modulation masking with sinusoidal carriers at high frequencies, broadband noise has generally been used as the carrier. This implies a broad excitation along the basilar membrane. The use of broadband-noise carriers, however, precludes investigation of temporal processing in the different frequency regions.

Chapter 2 of this thesis deals with narrowband carriers at a high center frequency whose bandwidth is chosen to be smaller than the bandwidth of the excited peripheral filter. Experiments on modulation detection and modulation masking are described which investigate the hypothesis of modulation channels. On the basis of these experiments a model based on one peripheral frequency channel is developed, incorporating a modulation filterbank whose parameters are adjusted so as to account for the experimental data. Results are discussed in terms of the statistical properties of the stimuli at the output of the excited modulation filters. The performance of the modulation filterbank model is compared with results from simulations obtained with a classical model (Viemeister, 1979).

Chapter 3 deals with spectral and temporal integration in amplitude modulation detection. It describes human performance at the transition of stimulus bandwidths within and beyond a critical bandwidth, and for broadband conditions. A multi-channel model is proposed to analyze the envelope fluctuations in parallel in each excited peripheral filter. The parameters of the modulation channels are assumed to be independent of frequency region, and the combination of information across frequency, i.e., the effect of spectral integration, is realized with the assumption of “independent” observations at the outputs of the different peripheral channels. Temporal integration refers to the ability of the auditory system to combine information over time to enhance the detection or discrimination of stimuli. It is important to distinguish between temporal resolution (or acuity) and temporal integration (or summation). The distinction between

these two “complementary” phenomena of resolution and integration does not necessarily mean that there must be two complementary modeling strategies to account for the data as proposed, for example, by Green (1985). Instead, the decision mechanism used in the present model (in combination with the preprocessing stages) is intended to allow a description of both the effects of temporal resolution (with time constants in the range of several milliseconds) and the effects of integration (with “effective” time constants in the range of hundreds of milliseconds).

Chapter 4 describes experiments on modulation detection using sinusoids at different carrier frequencies (in the range from 2–9 kHz). The assumption of independent observations across frequency made above is valid for random noise carriers. In such a case, the information about the presence of a signal modulation increases with the number of independent channels. However, the situation might be more complicated in conditions with deterministic carriers (such as sinusoids). Modulation thresholds can no longer be determined by the statistics of the inherent fluctuations of the stimuli, as in the conditions of the first two chapters. In the framework of the present model, performance should be solely limited by the variance of the *internal* noise, introduced at the end of the preprocessing stages. The detection of amplitude modulation in the range from 10–800 Hz is measured and compared with simulated thresholds obtained with the modulation- filterbank model. The tested conditions include the transition from purely temporal cues, such as roughness and loudness changes (at low modulation rates), to spectral cues (at high modulation rates), when the sidebands of the modulated stimuli are resolved by the auditory system.

Chapter 2

Amplitude modulation detection and masking with narrowband carriers¹

Abstract

*This paper presents a quantitative model for describing data from modulation-detection and modulation-masking experiments, which extends the model of the “effective” signal processing of the auditory system described in Dau et al. [J. Acoust. Soc. Am. **99**, 3615–3622 (1996a)]. The new element in the present model is a modulation filterbank, which exhibits two domains with different scaling. In the range 0–10 Hz, the modulation filters have a constant bandwidth of 5 Hz. Between 10 Hz and 1000 Hz a logarithmic scaling with a constant Q -value of 2 was assumed. To preclude spectral effects in temporal processing, measurements and corresponding simulations were performed with stochastic narrowband-noise carriers at a high center frequency (5 kHz). For conditions in which the modulation rate (f_{mod}) was smaller than half the bandwidth of the carrier (Δf), the model accounts for the lowpass characteristic in the threshold functions [e.g. Viemeister, J. Acoust. Soc. Am. **66**, 1364–1380 (1979)]. In conditions with $f_{mod} > \frac{\Delta f}{2}$, the model can account for the highpass characteristic in the threshold function. In a further experiment, a classical masking paradigm for investigating frequency selectivity was adopted and translated to the modulation-frequency domain. Masked thresholds for sinusoidal test modulation in the presence of a competing modulation masker were measured and simulated as a function of the test modulation rate. In all cases, the model describes the experimental data to within a few dB. It is proposed that the typical low-pass characteristic of the temporal modulation*

¹Modified version of the paper “Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers”, written together with Birger Kollmeier and Armin Kohlrausch, submitted to J. Acoust. Soc. Am.

transfer function observed with wideband noise carriers is not due to “sluggishness” in the auditory system, but can instead be accounted for by the interaction between modulation filters and the inherent fluctuations in the carrier.

2.1 Introduction

Periodic envelope fluctuations are a common feature of acoustic communication signals. The temporal features of vowel-like sounds, for example, can be described by a series of spectral components with a common fundamental frequency. Since the human cochlea has a limited frequency resolution, the higher frequency components are processed together in one frequency channel, that is, they stimulate the same group of hair cells and therefore are not separated spectrally within the auditory system. Two adjacent components of a harmonic sound which fall into the same frequency channel produce a form of amplitude modulation with a frequency corresponding to their difference frequency, which is equal to the fundamental frequency of the harmonic sound. In this way the fundamental can be encoded within that specific frequency channel, although it is physically absent. The “disadvantage” of the poor spectral resolution of simultaneously presented frequencies is thus compensated for by the “advantage” of temporal interaction between the spectrally unresolved components. Therefore, the temporal features of vowel-like sounds are in principle comparable to and are coded in a similar way to those of amplitude-modulated tones. The spectral peaks of the speech signal - the formants - would be considered as the carrier frequencies of amplitude modulations and the fundamental frequency of the vowel would correspond to the modulation frequency (Langner, 1992).

Temporal resolution of the auditory system, that is the ability to resolve dynamic acoustical cues, is very important for the processing of complex sounds. A general psychoacoustical approach to describing temporal resolution is to measure the threshold for detecting changes in the amplitude of a sound as a function of the rate of the changes. The function which relates threshold to modulation rate is called the temporal modulation transfer function (TMTF) (Viemeister, 1979). The TMTF might provide important information about the processing of temporal envelopes. It is often referred to as the time-domain equivalent of the audiogram, since it shows the “absolute” threshold for an amplitude-modulated waveform as a function of the modulation frequency. Since the modulation of a sound modifies its spectrum, wideband noise is often used as a carrier signal in order to prevent subjects using changes in the overall spectrum as a detection cue; modulation of white noise does not change its long-term spectrum. The subject’s sensitivity for detecting sinusoidal amplitude modulation of a broadband noise carrier is high for low modulation rates and decreases at high modulation rates. It is therefore often argued in the literature that the auditory system is too “sluggish” to follow fast temporal envelope fluctuations of sound. Since this sensitivity to modulation resembles the transfer function of a simple lowpass filter, the attenuation characteristic is often interpreted as the *lowpass characteristic* of the auditory system. This view is reflected in the structure of a popular model for describing the TMTF (Viemeister, 1979).

Measurements of the TMTF were initially motivated by the idea that tem-

poral resolution could be modeled using a linear systems approach (Viemeister, 1979). In a linear system the response to any input stimulus can be predicted by summing the responses to the individual sinusoidal components of that stimulus. A *time constant* is often derived from the modulation detection data - as the conjugate Fourier variable of the TMTF's cut-off frequency - to obtain an estimate of temporal acuity.

It is often argued that the auditory filters play a role in limiting temporal resolution (e.g., Moore and Glasberg, 1986), especially at low frequencies (below 1 kHz) where the bandwidths of the auditory filters are relatively narrow, leading to longer impulse responses (“ringing” of the filters). However, the response of auditory filters at high frequencies is too fast to be a limiting factor in most tasks of temporal resolution. Thus there must be a process at a level of the auditory system higher than the auditory nerve which limits temporal resolution and causes the “sluggishness” in following fast modulations of the sound envelope.

Results from several studies concerning modulation *masking*, however, are *not* consistent with the idea of only one broad filter, reflected in the TMTF. Modulation masking provides insight into how the auditory system processes temporal envelopes in the presence of another competing, temporally fluctuating background sound. Houtgast (1989) designed experiments to estimate the degree of frequency selectivity in the perception of simultaneously presented amplitude modulations, using broadband noise as a carrier. He adopted the classical masking paradigm for investigating frequency selectivity: the subject's task was to detect a test modulation in the presence of a masker modulation, as a function of the frequency difference between the two modulations rates. Houtgast found some correspondence with classical data on frequency selectivity in the audio-frequency domain. Using narrow bands of noise as the masker modulation, the modulation detection threshold function showed a peak at the masker modulation frequency. This indicates that masking is most effective when the test modulation frequency falls within the masker-modulation band. In the same vein, Bacon and Grantham (1989) found peaked masking patterns using sinusoidal masker modulation instead of a noise-band. Fassel (1994) found similar masking patterns using sinusoids at high frequencies as carriers and sinusoidal masker modulation.

For spectral tone-on-tone masking, effects of frequency selectivity are well established and associated with the existence of independent frequency channels (critical bands). When translated to the modulation frequency domain, the data of Houtgast, and Bacon and Grantham suggest the existence of modulation frequency specific channels at a higher level in the auditory pathway. Yost et al. (1989) also suggested amplitude modulation channels to explain their modulation detection interference (MDI) data and to account for the formation of auditory “objects” based upon common modulation. Martens (1982) had already suggested that the auditory system realizes some kind of short-term spectral analysis of the temporal waveform of the signal's envelope.

Modulation-frequency specificity has also been observed in different *physio-*

logical studies of neural responses to amplitude modulated tones (Creutzfeldt et al., 1980, Langner and Schreiner, 1988; Schreiner and Urbas, 1988). Langner (1992) summarized current knowledge about the representation and processing of periodic signals, from the cochlea to the cortex in mammals. Langner and Schreiner (1988) stated that the auditory system contains several levels of systematic topographical organization with respect to the response characteristics that convey temporal modulation aspects of the input signal. They found that these different levels of organization range from a general trend of changes in the temporal resolution along the ascending auditory axis (with a deterioration of resolution towards higher stations) to a highly systematically organized *map* of best modulation frequencies (BMF) within the inferior colliculus of the cat. Langner and Schreiner (1988) concluded that temporal aspects of a stimulus, such as envelope variations, represent a further major organizational principle of the auditory system, in addition to the well-established spectral (tonotopic) and binaural organization.

Of course, it is very difficult to establish functional connections between morphological structures and perception (cf. Viemeister and Plack, 1993; Schreiner and Langner, 1988; Fastl, 1990), and, furthermore, it is problematic to extrapolate from one species to another. In this sense, *psychophysics* may be the only presently available way to explore what mechanisms are needed, because it measures the *whole* nervous system in normal operation, and is not just concerned with specific neural activity, but with complex perception (Kay, 1982). On the other hand there is a boundless variety of mechanisms that could be postulated on the basis of psychoacoustical experiments. Given these difficulties, it would seem preferable to keep modeling within physiologically realistic limits.

The present psychoacoustical study further analyzes the processing of amplitude modulation in the auditory system. The goal is to gather more information about modulation frequency selectivity and to set up corresponding simulations with an extended version of a model of the "effective" signal processing in the auditory system, which was initially developed to describe masking effects for simultaneous and nonsimultaneous masking conditions and which is extensively described in Dau (1992) and Dau et al. (1995a,b). As already pointed out, in most classical studies about temporal processing a broadband noise carrier has been applied to determine the TMTF. This has the advantage that, in general, no spectral cues should be available to the subject, because the long-term spectrum of sinusoidally amplitude modulated noise (SAM noise) is flat and invariant with changes in modulation frequency. It is assumed that in general short-term spectral cues are not being used by the subject (Viemeister, 1979; Burns and Viemeister, 1981). On the other hand, as a great disadvantage, the use of broadband noise carriers does *not* allow direct information about spectral effects in temporal processing. Broadband noise excites a wide region of the basilar membrane, leaving unanswered the question of what spectral region or regions are being used to detect the modulation.

Therefore measurements and corresponding simulations with stochastic *narrowband* noises as the carrier at a *high* center frequency were performed, as was done earlier by Fleischer (1982). At high center frequencies the bandwidth of the auditory filters is relatively large so that there is a larger frequency range over which the sidebands resulting from the modulation are not resolved. Rather the modulation is perceived as a temporal attribute like fluctuations in loudness (for low modulation rates) or as roughness (for higher modulation rates). The bandwidth of the modulated signal is chosen in order to be smaller than the bandwidth of the stimulated peripheral filter. This implies that all spectral components are processed together and that temporal effects are dominant over spectral effects.

2.2 Description of the model

2.2.1 Original model of the “effective” signal processing

In Dau (1992), Dau and Püschel (1993) and Dau et al. (1995a,b) a model was proposed to describe the “effective” signal processing in the auditory system. This model allows the prediction of masked thresholds in a variety of simultaneous and non-simultaneous conditions. The model was initially designed to describe temporal aspects of masking. There is no restriction as to the duration, spectral composition and statistical properties of the masker and the signal.

The model combines several stages of preprocessing with a decision device that has the properties of an optimal detector. Figure 2.1 shows how the different processing stages in the auditory system are realized in the model. The frequency-place transformation on the basilar membrane is simulated by a linear basilar-membrane model (Schroeder, 1973; Strube, 1985). Only the channel tuned to the signal frequency is further examined. As long as broadband noise maskers are used, the use of off-frequency information is not advantageous for the subjects. The signal at the output of the specific basilar-membrane segment is half-wave rectified and lowpass filtered at 1 kHz. This stage roughly simulates the transformation of the mechanical oscillations of the basilar membrane into receptor potentials in the inner hair cells. The lowpass filtering essentially preserves the envelope of the signal for high carrier frequencies.

Effects of adaptation are simulated by feedback loops (Püschel, 1988; Kohlrausch et al., 1992). The model tries to incorporate the adaptive properties of the auditory periphery. It was initially developed to describe forward masking data. Adaptation refers to dynamic changes in the transfer gain of a system in response to changes in the input level. The adaptation stage consists of a chain of five feedback loops in series, with different time constants. Within each single element, the lowpass filtered output is fed back to form the denominator of the dividing element. The divisor is the momentary charging state of the lowpass filter, determining the attenuation applied to the input. The time constants range from 5 to 500 ms. In a stationary condition, the output of each element is equal to the square root of the input. Due to the combination of five elements the *stationary* transformation has a compression characteristic which is close to the *logarithm* of the input. Fast fluctuations of the input are transformed more linearly (see also section 2.2.2.1). In the stage following the feedback loops, the signal is lowpass filtered with a time constant of 20 ms, corresponding to a cutoff frequency of nearly 8 Hz to account for effects of temporal integration.

To model the limits of resolution an internal noise with a constant variance is added to the output of the preprocessing stages. The transformed signal after the addition of noise is called the *internal representation* of the signal. The auditory signal processing stages are followed by an optimal detector whose performance is limited by the nonlinear processing and the internal noise. The main idea of

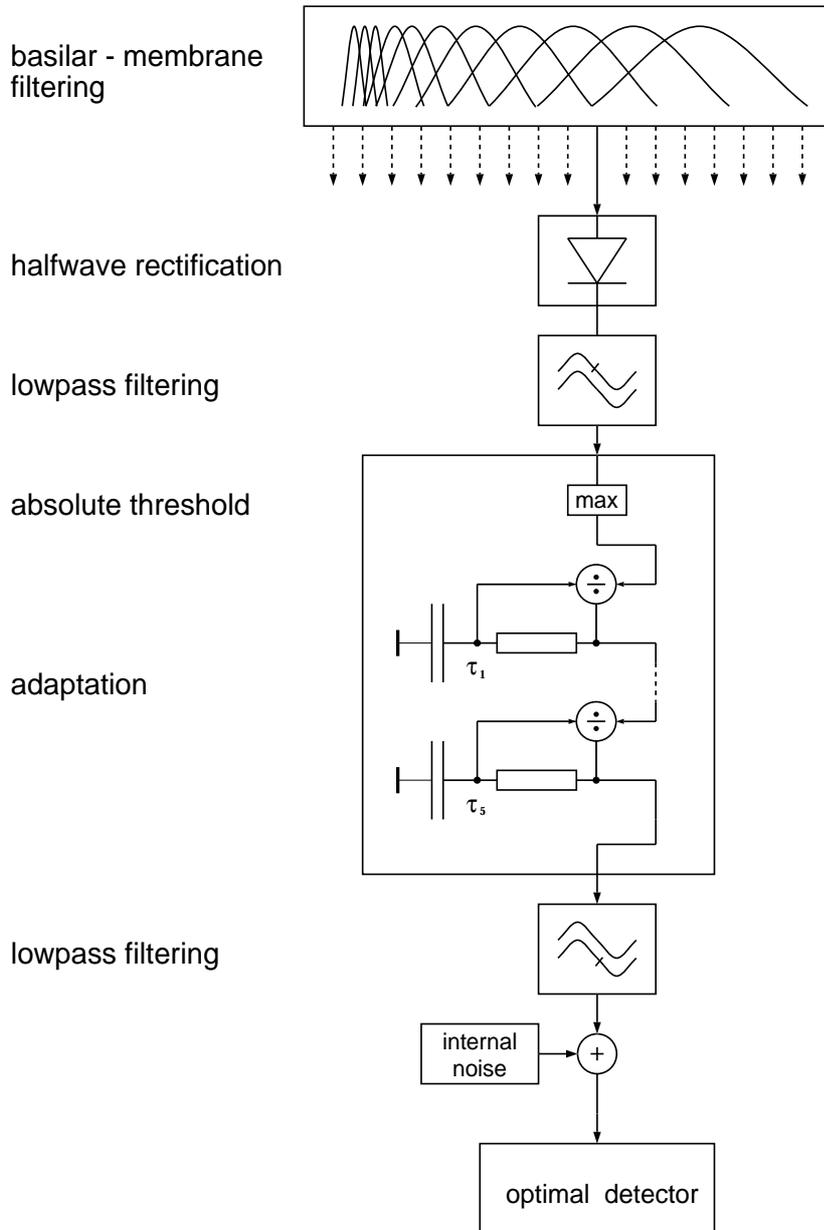


Figure 2.1: Block diagram of the psychoacoustical model for describing simultaneous and nonsimultaneous masking data with an optimal detector as decision device (Dau, 1992; Dau et al., 1995a). The signals are preprocessed, fed through nonlinear adaptation circuits, lowpass filtered and finally added to internal noise; this processing transforms the signals into their internal representations.

the optimal detector is that a change in a test stimulus is just detectable if the corresponding change in the internal representation of that test stimulus - compared with an internally stored reference - is large enough to emerge significantly from the internal noise. In the decision process, a stored temporal representation of the signal to be detected (the template) is compared with the actual activity pattern evoked on a given trial. The comparison amounts to calculating the cross correlation between the two temporal patterns and is comparable to a “matched filtering” process. The detector itself derives the template at the beginning of each simulated threshold measurement from a suprathreshold value of the stimulus. If signals are presented using the same type of adaptive procedure as in corresponding psychoacoustical measurements, the model could be considered as “imitating” a human observer. The optimality of the detection process refers to the best possible theoretical performance in detecting signals under specific conditions. The details about the optimal detection stage using signal detection theory (Green and Swets, 1966) are described in Appendix A. The calibration of the model is based on the 1-dB criterion in intensity discrimination tasks. In the first step of adjusting the model parameters, this value of a just-noticeable change in level of 1 dB was used to determine the variance of the internal noise.

In the model described above, the stimulus - in its representation after the adaptation stage - is filtered with a time constant of 20 ms. This stage represents the “hard-wired” integrative properties of the model and leads - in combination with preprocessing and the decision device - to very good agreement between experimental and simulated masked-threshold data. However, for describing modulation detection data it is not reasonable to limit the availability of information about fast temporal fluctuations of the envelope in that way. In addition, as pointed out in the Introduction, results from several studies concerning modulation masking indicate that there is some degree of frequency selectivity for modulation frequency. It is assumed here that the auditory system realizes some kind of spectral decomposition of the temporal envelope of the signals. For this reason, the following model structure is proposed to describe data on modulation perception.

2.2.2 Extension of the model for describing modulation perception

2.2.2.1 Stages of processing

Figure 2.2 shows the model that is proposed to describe experimental data on modulation perception. Instead of the implementation of the basilar-membrane model developed by Strube (1985) the gammatone filterbank model of Patterson et al. (1987) is used to simulate the bandpass characteristic of the basilar membrane. The parameters of this filterbank have been adjusted to fit psychoacoustical investigations of spectral masking using the notched-noise paradigm

(Patterson and Moore, 1986; Glasberg and Moore, 1990). The gammatone filterbank has the disadvantage that the phase characteristic of the transfer function of the basilar membrane is not described correctly, in contrast to the Strube model (Kohlrausch and Sander, 1995). For the experiments discussed in this paper, however, phase information plays a secondary role. Furthermore, in terms of computation time, the gammatone filterbank is much more efficient than the algorithm of the Strube model. The signal at the output of the specific filter of the gammatone filterbank is, as in the model described above, half-wave rectified and lowpass filtered at 1 kHz.

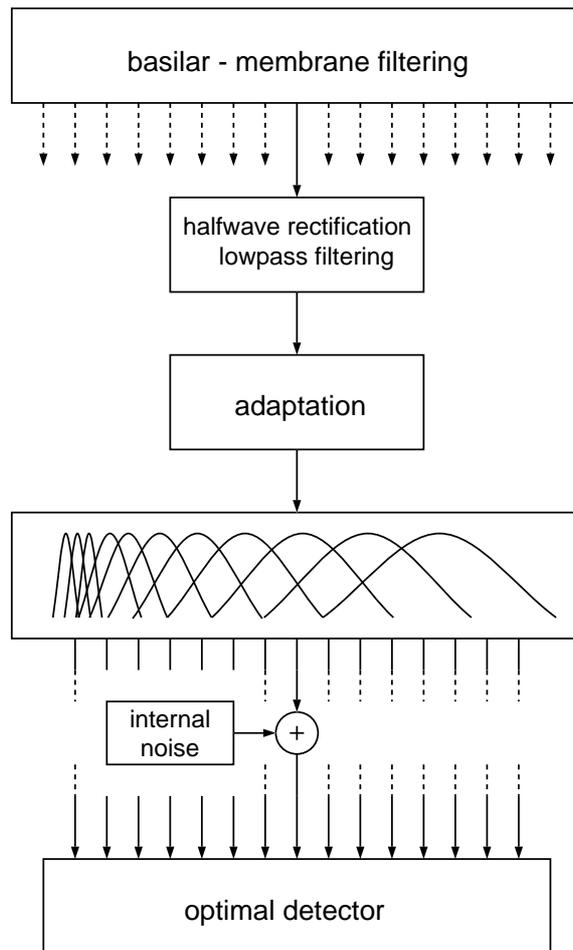


Figure 2.2: Block diagram of the psychoacoustical model for describing modulation detection data with an optimal detector as decision device. The signals are preprocessed, subjected to adaptation, filtered by a modulation filterbank and finally added to internal noise; this processing transforms the signals into their internal representations.

With regard to the transformation of envelope variations of the signal, the

nonlinear adaptation model (as implemented within the masking model) has the important feature that input variations that are rapid compared with the time constants of the lowpass filters are transformed *linearly*. If these changes are slow enough to be followed by the charging state of the capacitor, the attenuation gain is also changed. Each element within the adaptation model combines a static compressive nonlinearity with a higher sensitivity for fast temporal variations.

The following stage in the model, as shown in Fig. 2.2, contains the most substantial changes compared to the model described above. Instead of the lowpass filter, a linear filterbank is assumed to further analyze the amplitude changes of the envelope. This stage will be called *modulation filterbank* throughout this chapter. The implementation of this stage is in contrast to the signal processing within other models in the literature (e.g. Viemeister, 1979; Forrest and Green, 1987). The output of the “preprocessing” stages can now be interpreted as a three-dimensional, time-varying activity pattern. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation filter output. The calibration of the model is again based on the 1-dB criterion in intensity discrimination tasks. A long-duration signal with a fixed frequency and a level of 60 dB SPL was presented as input to the model. The variance of the internal noise was adjusted so that the adaptive procedure led to an increment threshold of approximately 1 dB. Because of the almost logarithmic compression of signal amplitude in the model, the 1-dB criterion is also approximately satisfied over the whole input level range. Because of the relatively broad tuning of the modulation filters (see section 2.2.2.2), some energy of the (stationary) signal also leaks into the transfer range of the overlapping modulation filters tuned to “higher” modulation frequencies. Therefore, a somewhat higher variance of the internal noise is required to satisfy the 1 dB-criterion compared to the variance adjusted with the modulation-lowpass approach described in the previous section.

The decision device is realized as an optimal detector in the same way as described in section 2.2.1 with the extension that in the present version the detector realizes a cross correlation between the three-dimensional internal representations of the template and the representation of the waveform on a given trial. The internal noises at the outputs of the different modulation channels are assumed to be *independent* from each other.

2.2.2.2 Modulation filterbank: Further model assumptions

It is often the case that models are developed to account only for a limited set of experiments or a single phenomenon. Each type of experiment leads to a model describing only the results of that experiment. As an example, de Boer (1985) considered several types of experiments on temporal discrimination: temporal integration, modulation detection and forward masking/gap detection and discussed the corresponding “ad hoc” models which cannot be united into one

model.

The present model tries to find a “link” between the description of phenomena of intensity discrimination and those of modulation discrimination. Assuming *linear* modulation filters analyzing the modulations of the incoming signals, the model would not be able to account for modulation masking data without any further nonlinearity. Masking means implicitly that there must be some kind of “information loss” at some level of auditory modulation processing.

To produce a loss of information in the processing of modulation, only the (Hilbert-) *envelope* of the different output signals of the modulation filterbank is further examined. This was suggested by Fassel (1994) to account for modulation masking data using a sinusoidal carrier. But what about the transformation

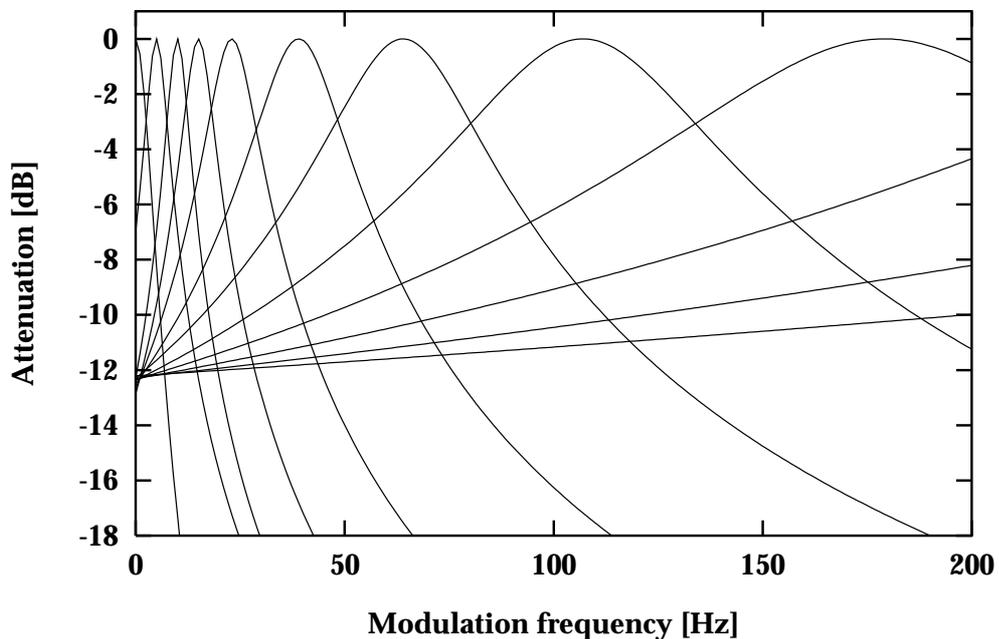


Figure 2.3: *Transfer functions of the modulation filters. In the range 0 – 10 Hz the functions have a constant bandwidth of 5 Hz. Above 10 Hz up to 1000 Hz a logarithmic scaling with a constant Q-value of 2 is applied. Only the range from 0-200 Hz is plotted.*

of very low modulation rates of the signal envelope? For these low rates it is not reasonable to extract the Hilbert envelope from the signal. It appears that the auditory system is very sensitive to slow modulations. Slow modulations are associated with the perception of rhythm. Samples of running speech, for example, show distributions of modulation frequencies with peaks around 3-4 Hz, approximately corresponding to the sequence rate of syllables (Plomp, 1983). Results from physiological studies have shown that, at least in mammals, the auditory cortex seems to be limited in its ability to follow fast temporal changes

in the input signal but, on the other hand, the cortex is capable of processing rhythm-like envelope fluctuations.

A further indication that use is made of information about modulation phase at low modulation rates is given by the successful description of intensity discrimination and masking data by the original model version described above, where the signal envelope was analyzed by the simple lowpass filter. This filtering preserves all information about the modulation phase.

It is postulated within the present model that the modulation filterbank exhibits two domains with different scaling. Figure 2.3 shows the transfer functions of the modulation filters. In the range 0 – 10 Hz a constant bandwidth of 5 Hz is assumed. From 10 Hz up to 1000 Hz a logarithmic scaling with a constant Q-value of 2 is assumed.² For all modulation channels with a center frequency above 10 Hz, only the envelope of the output signals is calculated.

The major part of this chapter describes and discusses experiments that have been designed to test the postulated “effective” signal processing. As a necessary and useful introduction to modulation experiments using narrowband noise carriers, the next section elaborates on some statistical properties of the envelopes of Gaussian noises. Although the mathematical roots of this section have already been published in 1950 by Lawson and Uhlenbeck, they have been ignored in most studies on modulation detection using Gaussian noise.

²The transfer functions of the resonance filters can be derived from the following recursive function: $y_i = e^{-i\pi B\Delta} \cdot e^{-i2\pi f_0\Delta} \cdot y_{i-1} + (1 - e^{-\pi B\Delta}) \cdot x_i$, where B is the filter bandwidth, f_0 is the center frequency of the resonance filter and Δ is the inverse samplerate.

2.3 Envelope statistics and envelope spectra of narrowband Gaussian noises

The envelope of a waveform, say $y(t)$, can be calculated from the waveform and its Hilbert transform:

$$z(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{y(t') dt'}{t' - t}. \quad (2.1)$$

The amplitude of the (linear) envelope is then defined by:

$$L(t) = \sqrt{[y(t)]^2 + [z(t)]^2}. \quad (2.2)$$

For the following derivations, it is assumed that $y(t)$ has a Gaussian distribution with mean zero and variance σ^2 . The Hilbert transform $z(t)$ has the same power spectrum and autocorrelation function as $y(t)$. Regarding $y(t)$ as the superposition of harmonic components with unrelated phases, the process of Hilbert transformation, which shifts the phase of every Fourier component by $\frac{\pi}{2}$, gives another but independent superposition of the same kind. Therefore, $y(t)$ and $z(t)$ represent two independent Gaussian distributions with the same parameters. The envelope is thus distributed according to a Rayleigh distribution (e.g., Bracewell, 1986).

$$p(L) = \frac{L}{\sigma^2} e^{-\frac{L^2}{2\sigma^2}} \quad (2.3)$$

with mean $\bar{L} = \sigma\sqrt{\frac{\pi}{2}}$ and variance $\overline{L^2} - (\bar{L})^2 = [2 - \frac{\pi}{2}]\sigma^2$.

Envelope power is defined as the square of the envelope, $P_{env}(t) = L(t)^2$. The time average of the envelope power is twice the average waveform power and therefore, it is independent of the noise bandwidth (cf., Hartmann and Pumplin, 1988). Therefore, for example, two Gaussian waveforms with the same power but with different bandwidths, have the same envelope power.

An interesting question is related to the *spectral* distribution of the envelope power. Lawson and Uhlenbeck (1950) calculated the spectrum of the envelope via Fourier transform of the autocorrelation function of $L(t)$. Assuming a rectangular shape of the power spectrum of the carrier, as shown in the left panel of Fig. 2.4, Lawson and Uhlenbeck have shown that the modulation spectrum $N = N(f_{mod})$, i.e., the power spectrum of the (linear) envelope of the carrier, is given approximately by the formula:

$$N_{\Delta f, \rho}(f_{mod}) \approx \pi \Delta f \rho \delta(f_{mod}) + \frac{\pi \rho}{4 \Delta f} (\Delta f - f_{mod}), \quad (2.4)$$

where Δf is the carrier bandwidth, ρ is the power spectral density and f_{mod} indicates modulation frequency. This modulation spectrum is plotted in the right panel of Fig. 2.4 (continuous line). Besides the dc-peak there is a triangular

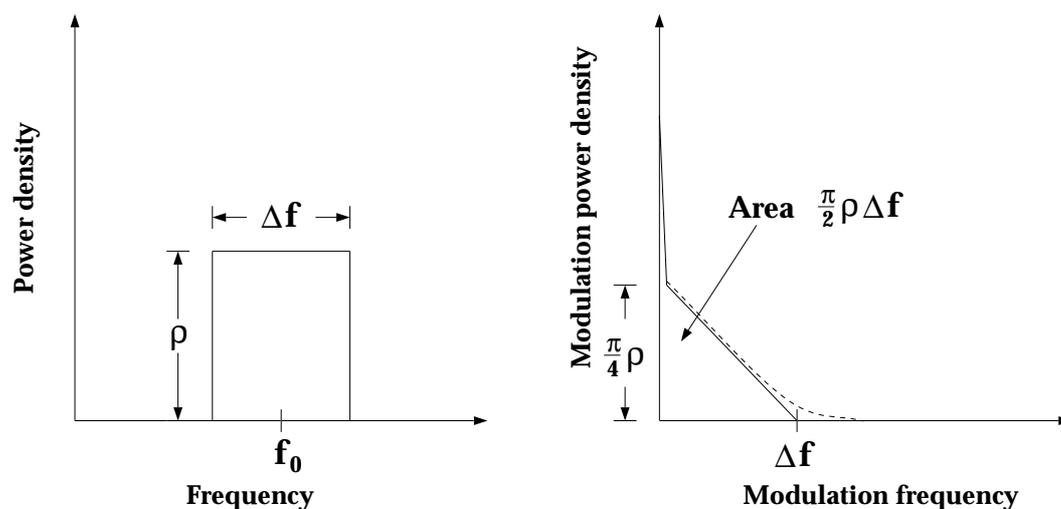


Figure 2.4: Power spectrum (left panel) and modulation spectrum (right panel) of bandlimited noise according to Lawson and Uhlenbeck (1950). The dashed curve shows the exact shape of the (linear) modulation spectrum, whereas the solid curve in the right panel represents an approximation.

continuous spectrum. This is a very good approximation. The exact shape is indicated by the dashed line in the same figure. In the case of the squared envelope, the modulation spectrum has *exactly* a triangular shape besides the de-peak.³ This corresponds to the Wiener-Chintchin theorem which states that the Fourier transform of the squared signal equals the autocorrelation of the spectrum of the signal.

The following aspects of Fig. 2.4 are of particular relevance for modulation detection experiments using narrow-band noise carriers. For a constant overall level of a noise band, the total power of intrinsic noise fluctuations, i.e., the total area under the triangle in the right panel of Fig. 2.4, remains constant. What changes is the spectral region over which the envelope spectrum stretches. With increasing noise bandwidth, the (modulation) spectrum becomes broader and flatter. For a low modulation frequency (typically below half the bandwidth of the noise band), this leads to a reduction in (intrinsic) modulation power. For higher modulation frequencies, above about half the noise bandwidth, the intrinsic modulation power increases with increasing noise bandwidth.

The statistical properties and, in particular, the spectral distribution of modulation power in the modulation spectrum have not been sufficiently taken into account in many studies on modulation perception. In fact, there are only a few studies in which the authors attempted to involve the *inherent* statistical

³According to Lawson and Uhlenbeck (1950), the modulation spectrum of the squared envelope, $Q(t) = x^2(t) + z^2(t)$, is given by the formula: $N_{\Delta f, \rho}^{square}(f_{mod}) = 8(\Delta f)^2 \rho^2 \delta(f_{mod}) + 8\rho^2(\Delta f - f_{mod})$.

properties of the noise carriers (e.g., Zwicker, 1953; Maiwald, 1967; Fleischer, 1981). In order to find a description of the inherent modulation of the noise, narrowband noise with a bandwidth Δf was regarded as a pure tone which was amplitude modulated by a continuum of equal-amplitude modulation frequencies between zero and half the bandwidth of the noise: $0 \leq f_{mod} \leq \frac{\Delta f}{2}$. To derive the total amount of intrinsic fluctuation power, two additional assumptions were introduced. First, the “modulation spectrum” was weighted by a factor $k_m(f_{mod})$, which essentially represented a low-pass characteristic. This factor was introduced to take into account the ear’s reduced ability to follow fast sound fluctuations compared to slow fluctuations (Fleischer, 1981 and 1982). Second, since the individual components in the modulation spectrum are incoherent, Fleischer assumed that they add according to their energies, so that the respective weighting factors have to be squared. After integration and normalization, the extraction of the square root led to the quantity:

$$k_n(\Delta f) = \sqrt{\frac{2}{\Delta f} \int_{-\infty}^{+\infty} k_m^2(f_{mod}) df_{mod}} \quad (2.5)$$

The *degree of fluctuation* of narrow band noise itself was defined as

$$g_n = k_n(\Delta f) \cdot m_n \quad (2.6)$$

where m_n is the effective degree of modulation of a noise signal excluding any spectral weighting of modulation frequencies.⁴

Furthermore, Fleischer (1982) assumed that, if an additional amplitude modulation is imposed on a narrowband noise, the total degree of fluctuation can be obtained by adding the respective “energies”, according to:

$$g = \sqrt{g_n^2 + g_m^2} = \sqrt{(k_n m_n)^2 + (k_m m)^2} \quad (2.7)$$

where g_n is the degree of fluctuation of the narrowband noise itself and g_m is the degree of fluctuation of the imposed amplitude modulation with a modulation degree m . Fleischer postulated that the auditory system detects differences between the modulations of successive acoustic signals as soon as the total degree of fluctuations increases by 20%. For more details, the reader is referred to Fleischer (1982). These considerations led to a prediction of a wide range of data on modulation perception. As Fleischer admitted, the complicated processes occurring when a modulated signal is modulated again, can be described by (2.7) only approximately.

⁴The depth of sinusoidal amplitude modulation is often characterized by the degree of modulation: $m = \frac{\hat{p}_{max} - \hat{p}_{min}}{\hat{p}_{max} + \hat{p}_{min}}$, based on the maximum \hat{p}_{max} and minimum \hat{p}_{min} of the sound pressure envelope. For narrow band noise, the value of $m_n = 0.4$ for the effective degree of modulation of noise signals gave the best fit to Fleischer’s data (Fleischer, 1982). Maiwald (1967) had worked with a value $m_n = 0.7$. The theoretical value based on the mathematical derivations in Lawson and Uhlenbeck is 0.5.

Further points of criticism concerning Fleischer's model can be discerned using the analytical results from Lawson and Uhlenbeck. The assumption that a narrow-band noise signal of bandwidth Δf is regarded as a tone which is modulated by a flat-spectrum continuum of modulation frequencies between zero and half the noise bandwidth, is not correct. It would imply that the modulation spectrum of noise has a flat shape in contrast to the triangular shape shown in Fig. 2.4. Finally, as will be shown later, it is problematic to model the observation that modulation thresholds increase with increasing modulation frequency by a weighting factor that has essentially a low pass characteristic.

2.4 Method

2.4.1 Procedure and Subjects

Modulation detection thresholds were measured and simulated using an adaptive 3-Interval Forced-Choice (3IFC) procedure. The carrier was presented in three consecutive intervals separated by silent intervals of 300 ms. In one randomly chosen interval the carrier was sinusoidally amplitude modulated. In the other intervals it was unmodulated. The subject's task was to specify the interval containing the modulation. During a threshold run, the modulation depth in dB ($20 \log m$), was adjusted using a 2-down 1-up rule (Levitt, 1971) which provides an estimate of the modulation depth necessary for 70.7% correct responses. The step size was 4 dB at the start of a run and was divided by 2 after every two reversals of the modulation depth until the step size reached a minimum of 1 dB, at which time it was fixed. Using this 1-dB step size, ten reversals were obtained and the median value of the modulation depths at these 10 reversals was used as the threshold value. The subjects received visual feedback during the measurements. The procedure was repeated four times for each signal configuration and subject. All figures show the median and interquartile ranges based on four single measurements. All five subjects had experience in psychoacoustic measurements and had clinically normal hearing. They were between 23 and 29 years old and participated voluntarily in the study.

2.4.2 Apparatus and stimuli

All acoustic stimuli were digitally generated at a sampling frequency of 30 kHz. The stimuli were transformed to analog signals with the aid of a two-channel 16-bit D/A converter, attenuated, lowpass filtered at 10 kHz and diotically presented via headphones (HDA 200) in a sound-attenuating booth. Signal generation and presentation were controlled by a SUN-Workstation using a signal-processing software package developed at the Drittes Physikalisches Institut in Göttingen.

Several modulation-detection and modulation-masking experiments were performed. In most measurements narrowband Gaussian noise centered at 5 kHz was used as the carrier. In the masking experiment a sinusoidal carrier at 5 kHz was used. The carrier level was 65 dB SPL in both cases. In the experiments using a noise carrier, an independent sample of noise was presented in each interval. The noise stimuli were digitally filtered before modulation by setting the magnitude of the Fourier coefficients to zero outside the desired passband.

In the case of the largest applied carrier bandwidth ($\Delta f = 314$ Hz), the stimuli were calculated differently: Broadband noise was first modulated, and then restricted in its bandwidth to 314 Hz. In this way possible spectral cues were avoided and the task was purely temporal in nature. This technique of introducing modulation prior to filtering has already been used by Viemeister

(1979) and Eddins (1993).

When generating amplitude-modulated narrowband stimuli, the average power of the modulated signal is increased by $1 + \frac{m^2}{2}$ compared with the unmodulated signal. For large modulation depths detection might therefore be based on changes in overall intensity rather than on the presence or absence of modulation. To eliminate level cues, the digital waveforms were adjusted to have equal power in each interval of the forced-choice trial.

In most cases sinusoidal test modulation was applied. In one experiment a complex modulator was used, consisting of five adjacent components of a harmonic tone complex. In each case the carrier and the applied modulators were windowed with a length depending on the particular experiment.

2.5 Results

2.5.1 Measurements and simulations of modulation detection and modulation masking

2.5.1.1 Amplitude modulation thresholds of narrowband noise as a function of the carrier bandwidth

Most of the following experiments are similar to those described by Fleischer (1982). He investigated amplitude modulation detection for noise carriers using a unmodulated carrier bandwidth of $\Delta f = 3, 31$ and 314 Hz at a sound pressure level of 80 dB. The subject's task was to detect the applied test modulation which ranged between 3 and 150 Hz. Fleischer defined the modulation threshold to be that modulation which yielded a proportion of correct detection of 50% . He obtained this modulation value from psychometric functions determined using a method of comparison in which the subjects had to note whether they perceived a difference between the modulated and unmodulated waveform.

Fleischer's experiments were replicated in this study and compared with corresponding simulations carried out with the present model. In contrast to Fleischer, an adaptive threshold procedure was used and the carrier level was somewhat lower (65 dB SPL). The carrier and the applied sinusoidal modulation had a duration of 1 s. Both were windowed with 200 ms cosine-squared ramps. Figure 2.5 shows the present experimental results for amplitude modulation detection employing a carrier bandwidth of 3 Hz at a center frequency of 5 kHz. The figure shows the data of three subjects (open symbols) together with the model predictions (closed symbols). The ordinate indicates modulation depth at threshold. The abscissa represents the modulation frequency. There is a very high detection threshold at a modulation rate of 3 Hz. This is due to the *inherent* statistical fluctuations of the narrowband 3 -Hz wide masker. The inherent fluctuations of the envelope mask the additional periodic 3 -Hz test modulation. With increasing modulation frequency thresholds decrease and converge with those obtained using a sinusoidal carrier at a modulation frequency of 20 Hz. These values for a sinusoidal carrier at 5 kHz are indicated by the asterisks. The threshold remains flat up to the modulation frequency of 100 Hz. This indicates that the auditory system does not seem too slow or sluggish to follow fast fluctuations (in this range), as is stated in many places in the literature. There is a very good agreement between measurements and simulations. Figure 2.6 shows thresholds using a narrowband carrier with a bandwidth of 31 Hz. Again, the modulation depth, m , at threshold is measured and simulated as a function of the test modulation frequency. The open symbols represent the measured data of three subjects and the filled symbols indicate the simulated thresholds. The threshold at a very low modulation rate (3 Hz) is several dB lower than in the case of the 3 -Hz wide carrier. This is due to the decreasing spectral density of the inherent

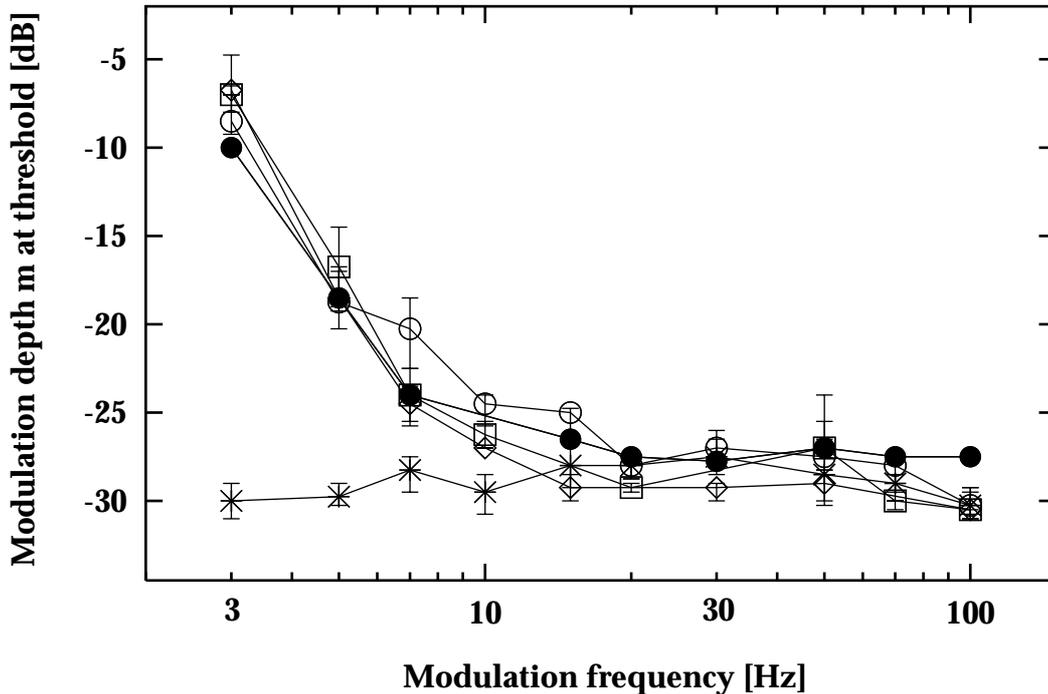


Figure 2.5: Modulation detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 3-Hz wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1s. Level: 65 dB SPL. Subjects: JV (□), AS (◇), TD (○), optimal detector (●). In addition, the modulation detection thresholds of one subject (TD) for a 5-kHz sinusoidal carrier is indicated by (★).

fluctuations with increasing bandwidth of the carrier, when the total energy of the modulated signal is constant (compare Sect. 2.3). In terms of the model, less “noise energy” falls into the low-frequency modulation filter which is tuned to the test modulation frequency. For modulation frequencies larger than half the bandwidth of the noise ($f_{mod} > \frac{\Delta f}{2}$) the threshold begins to decrease, in the measurements as well as in the simulations. That is, a highpass characteristic in the threshold function becomes apparent. However, thresholds decrease *more slowly* with increasing modulation frequency than the spectrum of the inherent envelope fluctuations itself (e.g., Lawson and Uhlenbeck, 1950). It is interesting in this context, that even at high modulation rates – at 100 Hz and 150 Hz – thresholds have not yet converged with those for the 3-Hz wide carrier nor with those for the sinusoidal carrier. The threshold is about 5 dB higher than with the 3-Hz wide carrier. This implies that the relatively slow inherent fluctuations of the 30-Hz wide carrier make it difficult to detect the higher-frequency test modulation. This phenomenon was also observed by Fleischer (1982) and was

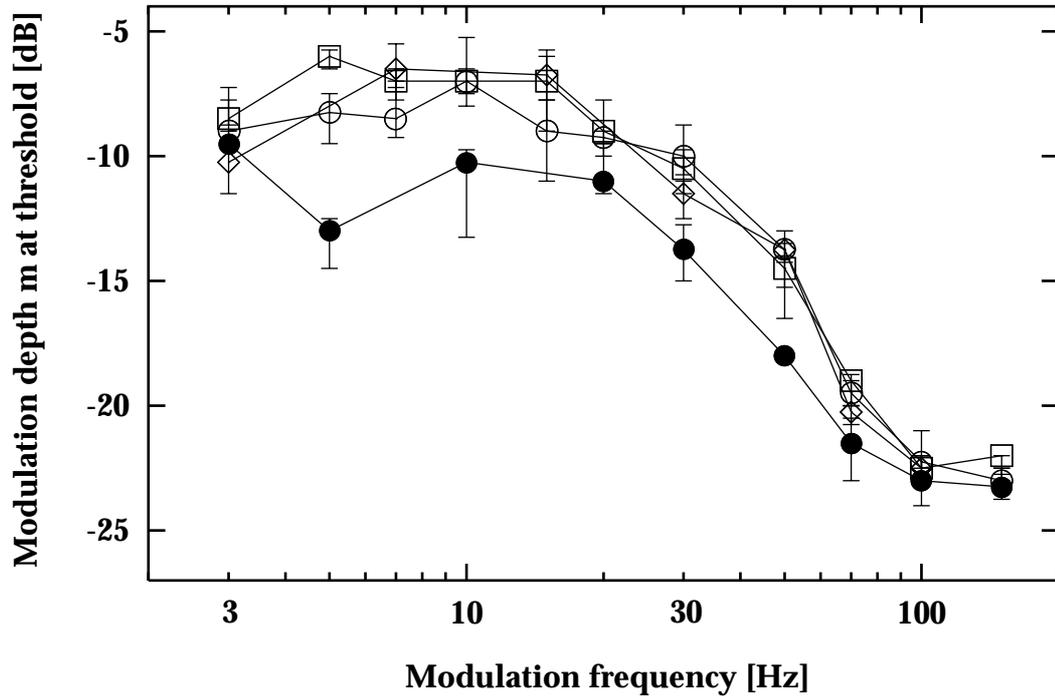


Figure 2.6: Modulation detection thresholds of a sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 31-Hz wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1s. Level: 65 dB SPL. Subjects: AS (◇), TD (○), JV (□), optimal detector (●).

denoted as “cross-talk” of the inherent fluctuations of the noise on the added modulation. This effect decreases with increasing rate of the test modulation.

This experiment reveals much about the auditory system’s selectivity for modulation frequency. In the model it was necessary to use wide modulation filters ($Q=2$) at high modulation frequencies so that some energy from the low-frequency fluctuations of the “masker” leaks through a modulation filter that is tuned to a high modulation frequency (like 150 Hz). This leakage decreases the signal-to-noise ratio and therefore leads to a higher detection threshold at $f_{mod} = 150$ Hz than would be the case for a more sharply tuned filter. Again, there is a good agreement between simulated and measured data.

Figure 2.7 shows results for the carrier bandwidth of 314 Hz. In this condition the bandwidth was digitally limited *after* modulation to eliminate spectral cues. The open symbols represent the measured data of three subjects whereas simulated thresholds are indicated by the filled symbols. The threshold is higher for a modulation rate of 3 Hz than for a rate of 5 Hz. This is probably caused by the use of a gated carrier. Such an effect of “adaptation” has been examined in several psychoacoustical and physiological studies (e.g., Viemeister, 1979; Sheft

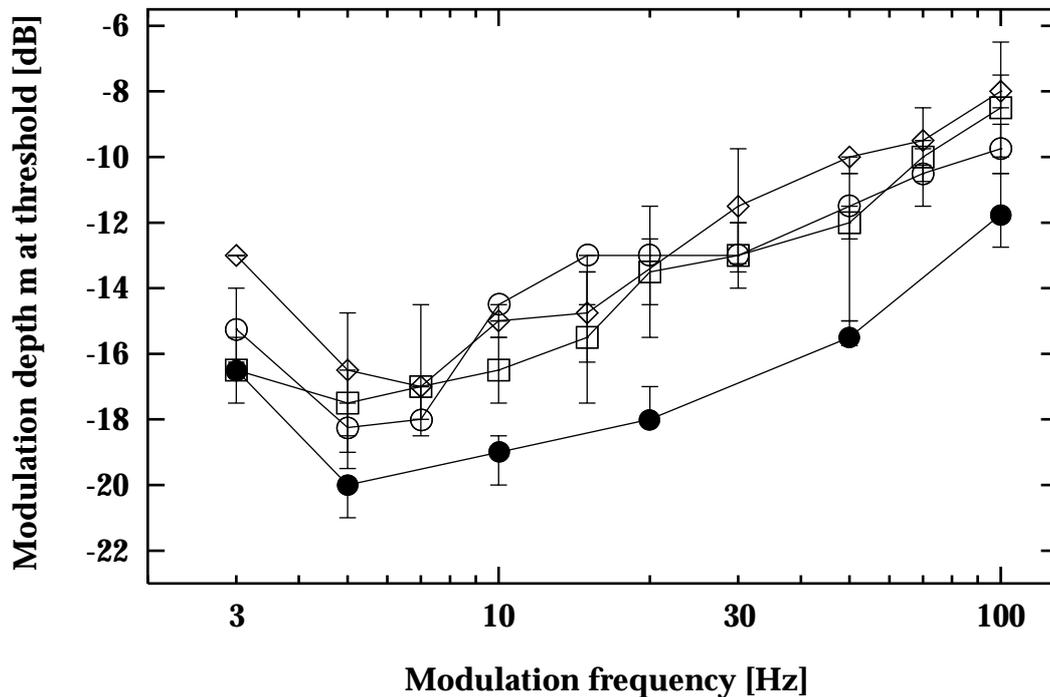


Figure 2.7: *Modulation detection thresholds of a sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 314-Hz wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1s. Level: 65 dB SPL. Subjects: Data from Fleischer (□), TD (○), AS (◇), optimal detector (●).*

and Yost, 1990). Based on their results it can be assumed that the threshold at 3 Hz would decrease if one used a continuous carrier instead of a gated one. For modulation frequencies above 7 Hz, the threshold increases by about 3 dB per doubling of the modulation frequency. This threshold pattern is similar to the pattern found in “classical” measurements of the TMTF using a broadband noise masker as a carrier, but it has a much lower cut-off frequency. Overall, the threshold curve is very different from those obtained with smaller carrier bandwidths. With increasing modulation frequency it now becomes harder to detect the test modulation. Also in the simulations, one obtains an increasing threshold curve with increasing modulation frequency. Using a carrier bandwidth of 314 Hz implies that all modulation filters (up to 150 Hz) are “stimulated” with the inherent random fluctuations of the carrier envelope. The additional test component falls in the passband of mainly one modulation filter. Assuming a constant decision criterion at threshold, the *logarithmic* scaling of the modulation filters with center frequencies above 10 Hz leads to a 3-dB increase of modulation depth, m , at threshold. In other words, to get the same signal-to-noise ratio at threshold,

a greater modulation depth is required with increasing modulation frequency.

Figure 2.8 gives an impression of how the signals are internally represented in the model, showing an example of how the template is derived from the internal representations of a suprathreshold test modulation and that of the unmodulated carrier alone. The upper panel shows the three-dimensional internal representation of a 3-Hz wide carrier (centered at 5 kHz). It represents the internal activity as a function of time and center frequency of the modulation filters. The ordinate is scaled in model units (MU). The modulation center frequencies range from 0 to 1000 Hz. As expected, since the total energy within the modulation spectrum of the signal is concentrated at very low modulation rates, only the lowest modulation filters are excited by the input signal, as indicated by the hatched lines in the figure. The middle panel of Fig. 2.8 shows the internal representation of the carrier, this time sinusoidally modulated with a modulation rate of 50 Hz at a highly detectable modulation depth. The test modulation mainly activates the modulation filter tuned to 50 Hz but also stimulates adjacent modulation filters because of the relatively low modulation-frequency selectivity assumed in the model. Again, the inherent fluctuations of the carrier itself activate the region at low modulation frequencies. However, because of the large spectral separation between the test modulation and the inherent fluctuations of the carrier, there is no interaction between the modulations, that is, no competing “noise” energy leaks into the transfer range of the test filter. As a consequence, the internal representation of the template which is derived by subtracting the upper panel from the middle one and normalizing the result, contains a representation of the temporal course of the test modulation without interfering frequencies from the carrier modulations. This template is plotted in the lower panel.

2.5.1.2 Amplitude modulation thresholds of third-octave-wide noise bands as a function of the center frequency

In the previous section it was observed that the detection threshold for amplitude modulation depends on the spectral density of the inherent fluctuations of the carrier, when the total energy of the modulated signal is constant. This was examined further in the following experiment using a third-octave wide noiseband as the carrier. The detection threshold for a modulation of 25 Hz was measured and simulated as a function of the center frequency of the band. In the model, only the output of the stimulated peripheral filter was analyzed. It was further assumed that the scaling of the modulation filters does not change with the peripheral frequency region. Figure 2.9 shows the modulation depth at threshold as a function of the center frequency of the third-octave band noise. The open symbols represent measured data of three subjects; simulated thresholds are shown by the filled circles. The measured modulation threshold decreases with increasing center frequency. The increasing absolute bandwidth causes a decreasing density of low-frequency inherent envelope fluctuations, if the total energy of the mod-

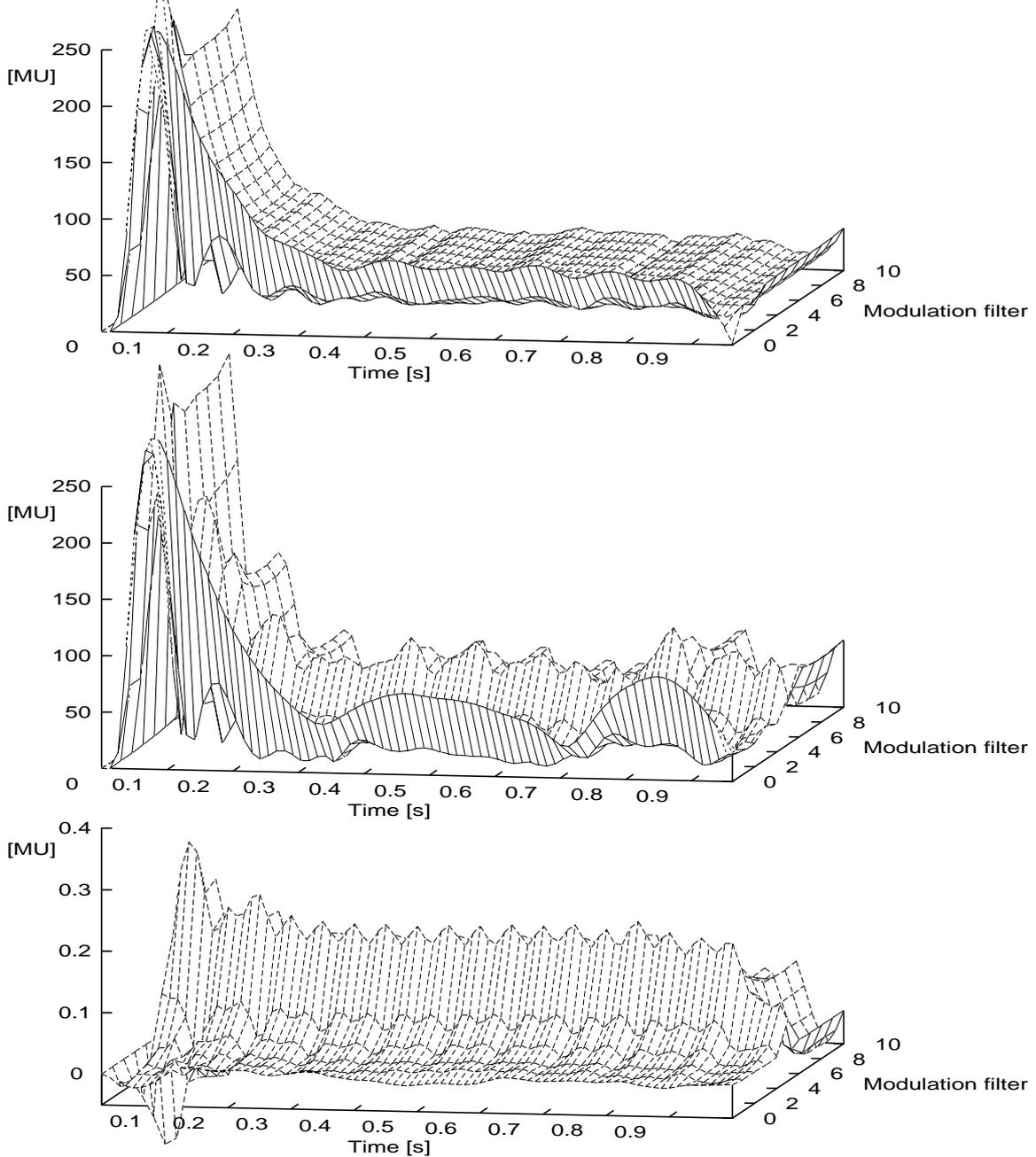


Figure 2.8: Generation of the template representation (at the bottom) of a 50-Hz test modulation which was impressed on a 3-Hz wide running-noise carrier centered at 5 kHz. The template is the normalized difference between the mean representation of the carrier plus the supra-threshold modulation (in the middle) and the mean representation of the carrier alone (at the top). The ordinate is scaled in model units (MU).

ulated stimulus is kept constant. As a consequence, less energy of the random envelope fluctuations of the carrier falls within the pass band of the modulation filter tuned to the test modulation frequency. This leads to decreasing thresholds with increasing center frequency in the model. Apart from the systematic 2-dB difference in the absolute sensitivity, there is good agreement between the simulated and measured data.

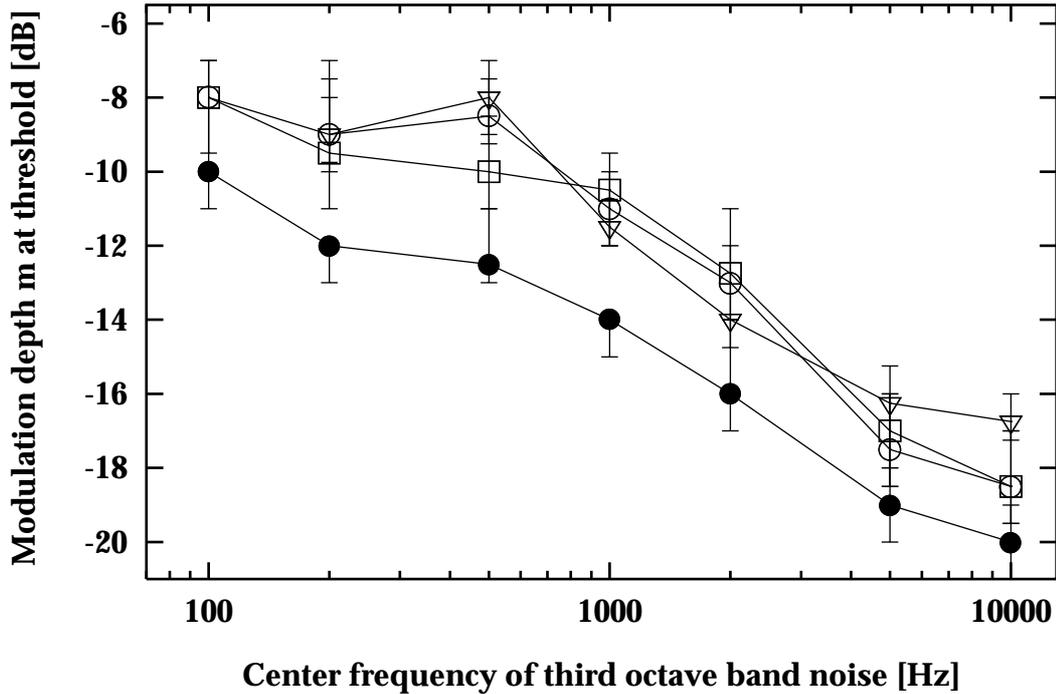


Figure 2.9: Modulation detection thresholds for 25-Hz modulation as a function of the center frequency of third-octave wide noise carrier. Carrier and modulation duration: 1s. Level: 65 dB SPL. Subjects: Data from Fleischer (1981) (□), TD (○), JV (▽), optimal detector (●).

2.5.1.3 Modulation masking: A harmonic tone complex as the masker

In a further experiment concerning modulation frequency selectivity, a masking paradigm for investigating frequency selectivity in the audio-frequency domain was adopted. It served as a test for spectral analysis in the modulation domain - as opposed to a periodicity analysis. A sinusoidal carrier at 5 kHz was used. As the masker modulation, a narrowband tone complex was used. This consisted of the third through seventh component of a harmonic tone complex with a fundamental frequency of 30 Hz, with frequencies 90, 120, 150, 180 and 210 Hz. In each interval, the starting phase of each spectral component was randomly

chosen from a uniform distribution in the range $0 - 360^\circ$. The amplitude of each component was 0.2 to guarantee that the resulting modulation depth of the tone complex never exceeded $m=1$. Because of the random phase, the modulating tone complex had a noise-like temporal waveform. As the signal, a sinusoidal test modulation was impressed on the same carrier. The test modulation was applied in the range between 20 and 120 Hz. Thus, the bandwidth of the modulated signal remained within the bandwidth of the auditory filter centered at 5 kHz. Figure 2.10 shows schematically the spectral distribution of masker and test components in the modulation spectrum. The modulated stimuli were pre-

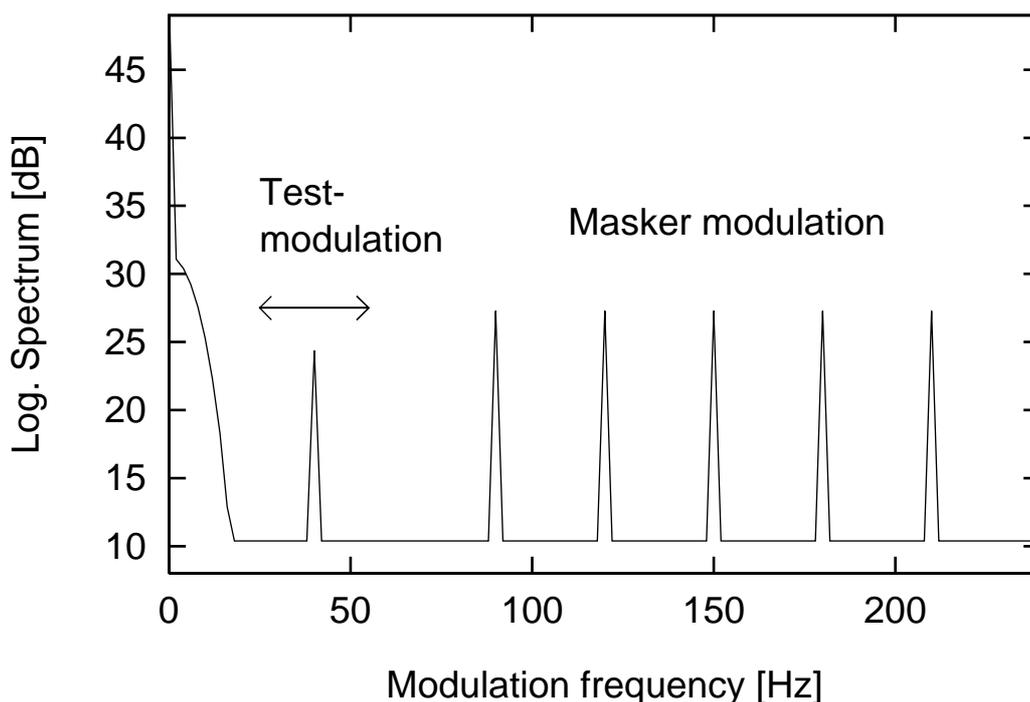


Figure 2.10: *Logarithmic spectrum of the Hilbert envelope of the stimuli presented in the signal interval. The signal interval contains the five components of the masking tone complex and the signal component. The subject's task was to detect the signal component.*

sented at a level of 65 dB SPL. They had a duration of 400 ms. Test and masker modulation were impressed for the whole duration of the carrier and were gated together with 20 ms Hanning ramps.

The amount of modulation masking as a function of the test modulation frequency is shown in Figure 2.11. The measured data are represented by the open symbols. The filled symbols indicate the simulated data. The “masking pattern” was derived by subtracting the unmasked threshold from the masked threshold at each test modulation frequency. The unmasked modulation thresholds, i.e. the threshold for a sinusoidal test modulation without any interfering masker modulation, were similar across subjects and were similar to those described in the

first experiment of this paper (see Fig. 4), where the stimuli were presented with a different duration. These unmasked thresholds were used as a reference to evaluate the effect of the modulated masker. The reference threshold (not plotted) remained more or less flat up to a modulation frequency of 120 Hz as already shown. As can be seen from Fig. 2.11, the amount of masking increased with

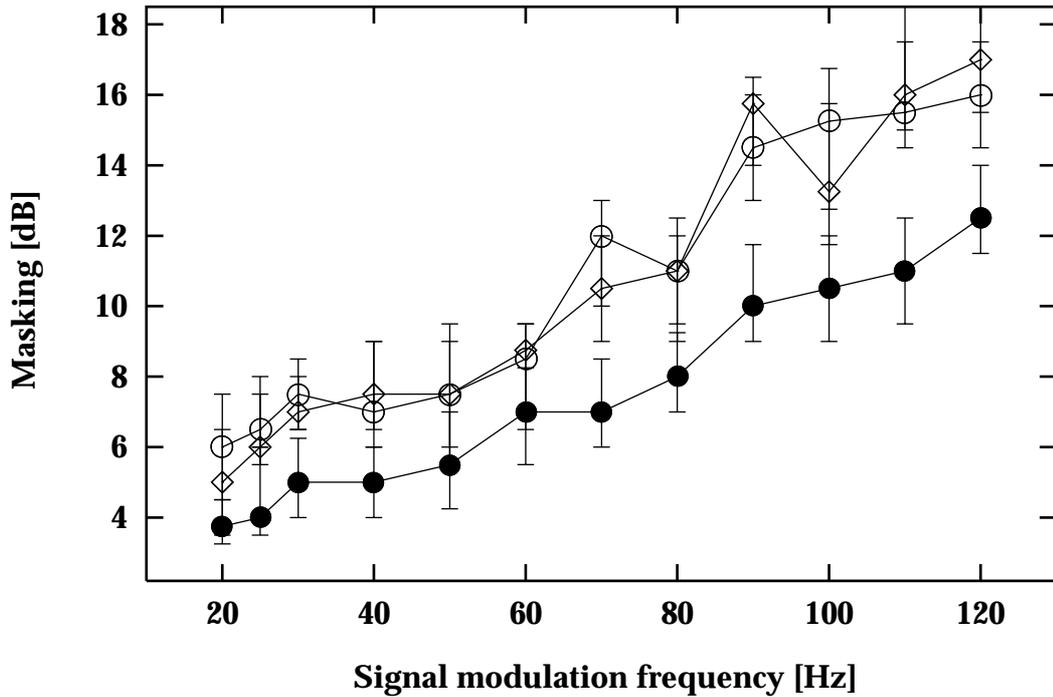


Figure 2.11: Amount of modulation masking as a function of the modulation frequency. Carrier: 5 kHz, Modulation masker: 3rd–7th components of a harmonic tone complex with fundamental $f_0 = 30$ Hz. Level: 65 dB SPL. Subjects: TD (○), JV (◇), optimal detector (●).

increasing test modulation frequency. The difference between the highest and the lowest threshold was more than 10 dB. It should be pointed out that there is no peak at 30 Hz (the “missing fundamental”). This indicates that the masking is not determined by the period of the modulation. An increasing masking effect with increasing test modulation frequency is also seen in the simulations. The masked threshold is directly related to the amount of masker energy falling within the passband of the modulation filter tuned to the actual test modulation. For the lowest test modulation rate (20 Hz) there is only a very small masking effect in the model since the modulation filters in the low modulation frequency region are assumed to be relatively sharply tuned (see Fig. 3). With increasing test modulation frequency, more and more components of the masker effectively

contribute to masking. Also, in the simulations the difference between the highest and the lowest masked threshold amounts to nearly 10 dB. However, there is a systematic difference of 2-4 dB between the measured and simulated masking patterns. However, it seems clear that there is no doubt that there is some degree of modulation frequency selectivity, even if this selectivity seems to be relatively broadly tuned.

2.5.2 Link between modulation detection and intensity discrimination

2.5.2.1 Phase effects at low modulation frequencies

It has been assumed within the present model that low modulation frequencies (0–10 Hz) are transformed without any information loss of the modulation phase, whereas for modulation frequencies above 10 Hz only the envelope of the output signals is further analyzed. To obtain further justification for this approach,

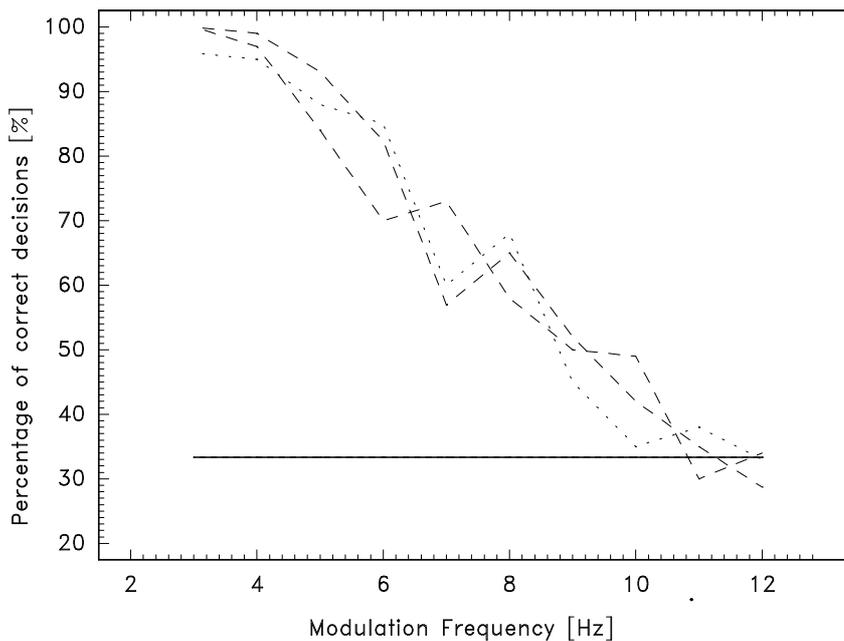


Figure 2.12: *Percentage of correct decisions ($P(C)$) in discriminating modulation phase as a function of the modulation frequency. Carrier: 5 kHz, modulation depth: $m = 1$, Level: 65 dB SPL. Subjects: TD, JV (dashed lines), AS (dotted line), chance probability (horizontal line).*

an experiment was performed to test the human sensitivity for the modulation phase at low modulation rates. For this, the experimental design was modified. A nonadaptive 3IFC-algorithm was used. The carrier was a 5-kHz sinusoid which

was sinusoidally modulated with $m = 1$ in each interval. The carrier had a duration of 800 ms and was windowed with Hanning-shaped ramps of 150 ms. The modulation had a duration of 500 ms and started 150 ms after carrier onset. The modulation was Hanning windowed over its entire duration. In each interval the same modulation frequency was impressed on the carrier. In the test interval the starting phase of the modulation was uniformly distributed between 0 and 360° from trial to trial. The reference intervals contained a constant difference in the starting phase ($\Delta\phi = 180^\circ$) of the modulation relative to the stimulus in the test interval. The subject's task was to determine the interval that contained the "differing" modulation starting phase (at fixed modulation depth). For each condition, i.e. for each applied modulation frequency, the percentage of correct decisions, $P(C)$, was derived from 100 successive trials and plotted as a function of the modulation frequency.

Figure 2.12 shows the data from three subjects. The ordinate represents the percentage of correct decisions ($P(C)$) and the abscissa represents the test modulation frequency. For all subjects, $P(C)$ decreases with increasing modulation frequency. While at the lowest modulation rate (3 Hz) the subjects' ability to discriminate the modulation starting phase is nearly perfect (100%), in contrast, $P(C)$ equals chance (33%) at a modulation rate of 12 Hz.

To test the role of the interaction between ramp and modulation during the rise and fall time of the modulation ramps the same experiment was carried out with the modification that the duration of the ramps was chosen to be two periods of the actual test modulation. In this way the interaction between modulation and ramp is comparable across the different conditions. The ability to discriminate modulation phase deteriorated with increasing test modulation frequency (not plotted), in the same way as before. Further experiments were performed to investigate modulation integration effects. No significant dependence of the results on the duration of the stationary part of the modulation in the range 300 to 700 ms was observed (not plotted).

It can be concluded that in monaural conditions using a sinusoid as the carrier, the auditory system's sensitivity to starting phase gets lost for modulation rates higher than around 12 Hz. Of course, in binaural conditions and also in monaural conditions where more complex carriers are presented, human performance is much better, indicating that across-channel information is used by the subjects. The simulation of such conditions, therefore, would require some across-channel processing. At the present time, this exceeds the goal of this study.

2.5.2.2 Experiments on test-tone integration and forward masking

The aim of this study was the development of a model that enables the prediction of both modulation detection and the more common signal detection data. As has been shown above, the implementation of the modulation filterbank allows the prediction of modulation discrimination data. Within this model, the signal

processing for very low modulation rates is nearly the same as in the previous version of the model (cf. Dau et al., 1995a). To explicitly check the ability of the new model to predict masking data as well as the previous model version, two simulations were performed. The first condition dealt with the phenomenon of temporal integration. In experiments with signals presented in noise maskers, the thresholds are found to decrease with increasing signal duration up to about 200 ms (e.g., Plomp and Bouman, 1959). The auditory system seems to be able to integrate signal intensity in detection tasks over at least 200 ms.

It was investigated whether the present model can account for this effect using a signal at 1 kHz presented in a 600-ms section of a broadband frozen-noise masker with a level of 77 dB SPL. Masked thresholds were determined using the same adaptive 3IFC-procedure and the same algorithm (2-down-1-up) as in the modulation tasks, with the exception that the signal level instead of the signal modulation depth was adjusted during the experimental run. Figure

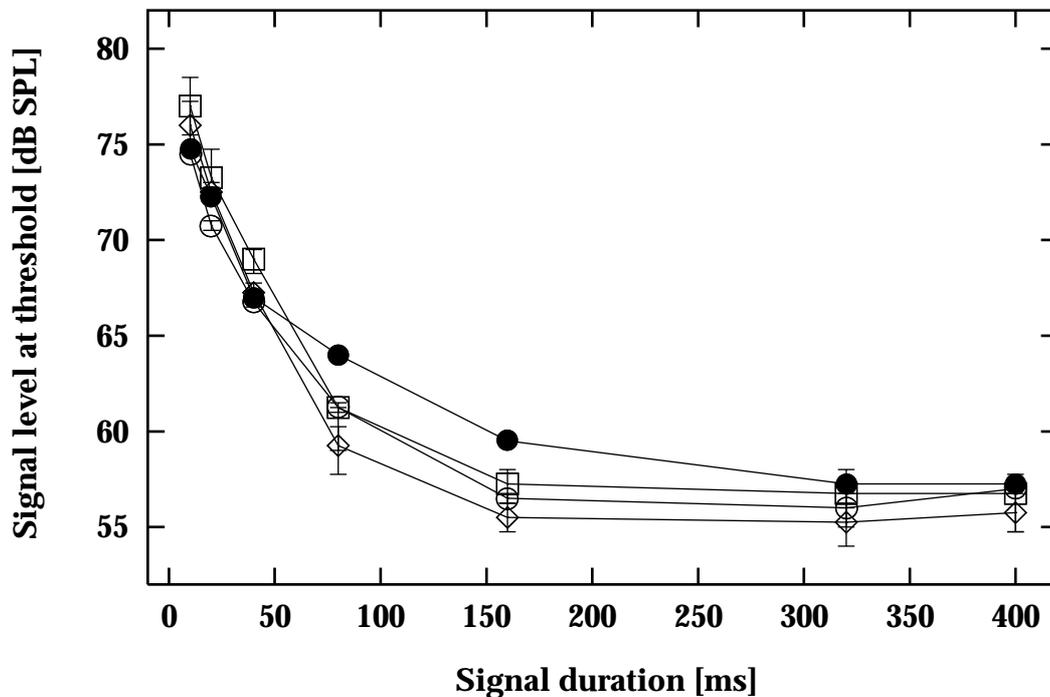


Figure 2.13: Masked thresholds of a signal presented in frozen noise for several signal durations. Signal: 1 kHz; Masker bandwidth: 20-5000 Hz, masker duration: 600 ms, masker level: 77 dB SPL. Subjects: AS (◇), AE (□), TD (○), optimal detector (●).

2.13 shows the results of the model (filled symbols) and those of three subjects (open symbols) for signal durations of 10, 20, 40, 80, 160, 320 and 400 ms. In the simulations only the channel tuned to the signal frequency was examined. The

experimental data are taken from Dau (1992). The onset for all signals occurred 100 ms after masker onset. The signal duration is indicated on the abscissa and the ordinate indicates the signal level at threshold. There is a 3 dB decay per doubling of signal duration in the measured data up to a duration of 160 ms. Such a decrease in the threshold function is also seen in the simulation. The agreement between measurements and simulations is very good and in fact better than that obtained with the previous model version (cf. Dau et al., 1995b).

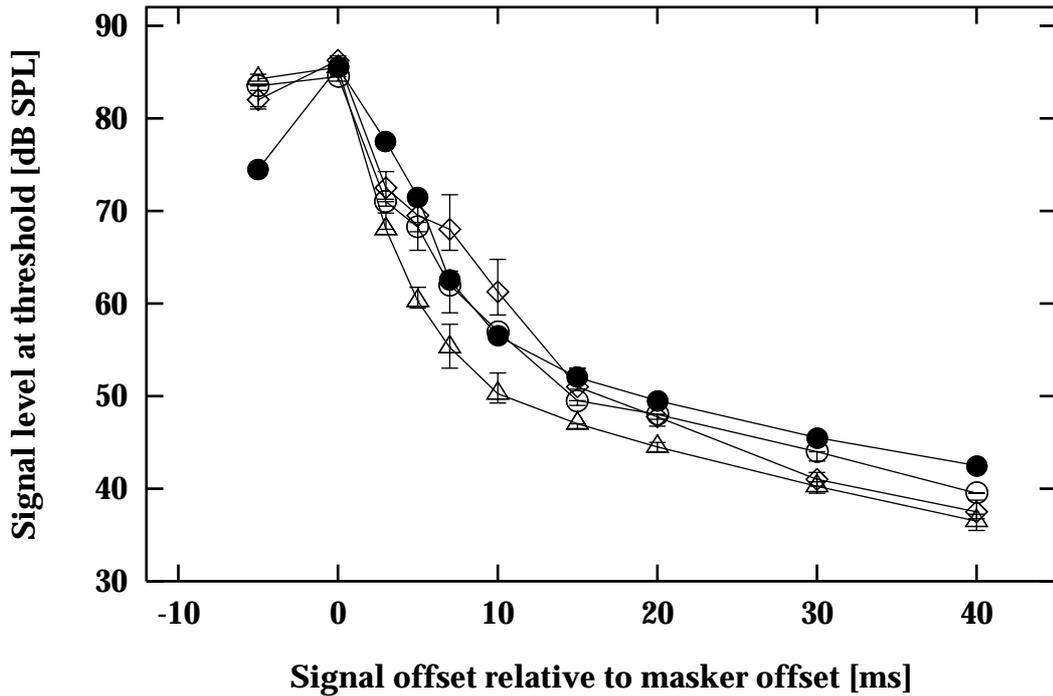


Figure 2.14: Forward-masked thresholds of a brief signal as a function of its temporal position relative to the offset of a 200-ms broadband noise masker. Signal: 5 ms, 2 kHz; Masker duration: 200 ms, masker bandwidth: 20-5000 Hz, masker level: 77 dB SPL. Subjects: AS (◊), SM (△), TD (○), optimal detector (●).

The second condition is concerned with forward masking. Such an experiment was initially carried out in order to test the implementation of the nonlinear adaptation stage (feedback loops). Within the earlier version of the model, the time constants of the nonlinear adaptation model were adjusted to fit the results of a “normal” listener. Since the results of the model calculations depend on the *whole* signal processing in combination with the decision device it was not clear in advance that the present model, in which the lowpass filter has been replaced by a modulation filterbank, can account for the data even if the parameters within the nonlinear adaptation stage are kept constant.

Figure 2.14 shows the simulated and measured thresholds for a 5-ms, 2-kHz signal presented at the end and after a fixed 200-ms section of a broadband frozen-noise masker. Thus, this forward masking condition includes the transition between simultaneous and forward masking. The figure shows the threshold level in forward masking as a function of the temporal separation between masker offset and signal offset. The filled symbols represent the simulated thresholds while the open circles indicate the measured data of three subjects, taken from Dau (1992). There is good agreement between simulated and measured data. The results of the model calculations hardly differ from those obtained with the lowpass filter approach (not shown) (see Dau et al., 1995b). This indicates that the most relevant information in intensity discrimination conditions is contained in the low frequency region of the modulation spectrum within the stimulated peripheral filter tuned to the signal frequency.

2.6 Predictions of Viemeister's model for modulation detection

The modulation filterbank concept differs considerably in its structure from the “classical” modulation lowpass filter approach (e.g., Viemeister, 1979). So far, it has been shown how well the modulation filterbank model describes experimental data. In this section, predictions of the modulation lowpass filter approach are investigated and compared with the performance of the modulation filterbank model.

The structure of Viemeister's model is shown in Fig. 2.15. The model incorporates a predetection bandpass filter (with a bandwidth of $\Delta f = 2000$ Hz) which is followed by a nonlinearity (half-wave rectification) and a lowpass filter. Viemeister fitted the cut-off frequency of the lowpass filter to the TMTF data obtained with a broadband noise carrier. The resulting cut-off frequency was 64 Hz. As decision variable he suggested the ac-coupled root-mean-square (rms) value of the output of the lowpass filter which was calculated over the duration of the observation interval. The thresholds were defined as the modulation depth necessary to produce a certain average increment in rms value, compared to that for an unmodulated noise.

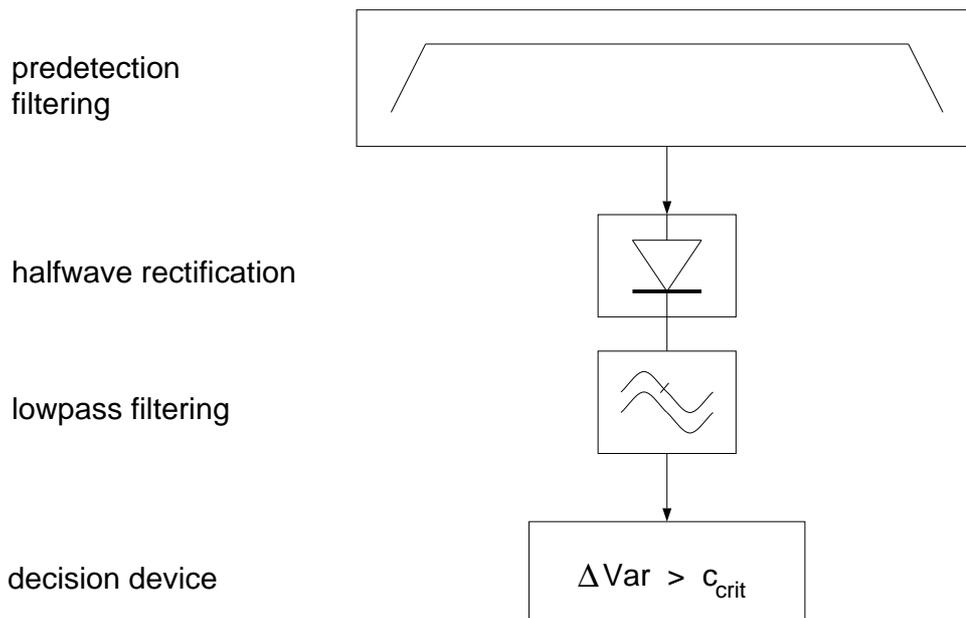


Figure 2.15: Model proposed by Viemeister (1979): It consists of a predetection bandpass filter with a bandwidth of 2000 Hz, followed by a half-wave rectification and a lowpass filter with a cut-off frequency of 64 Hz.

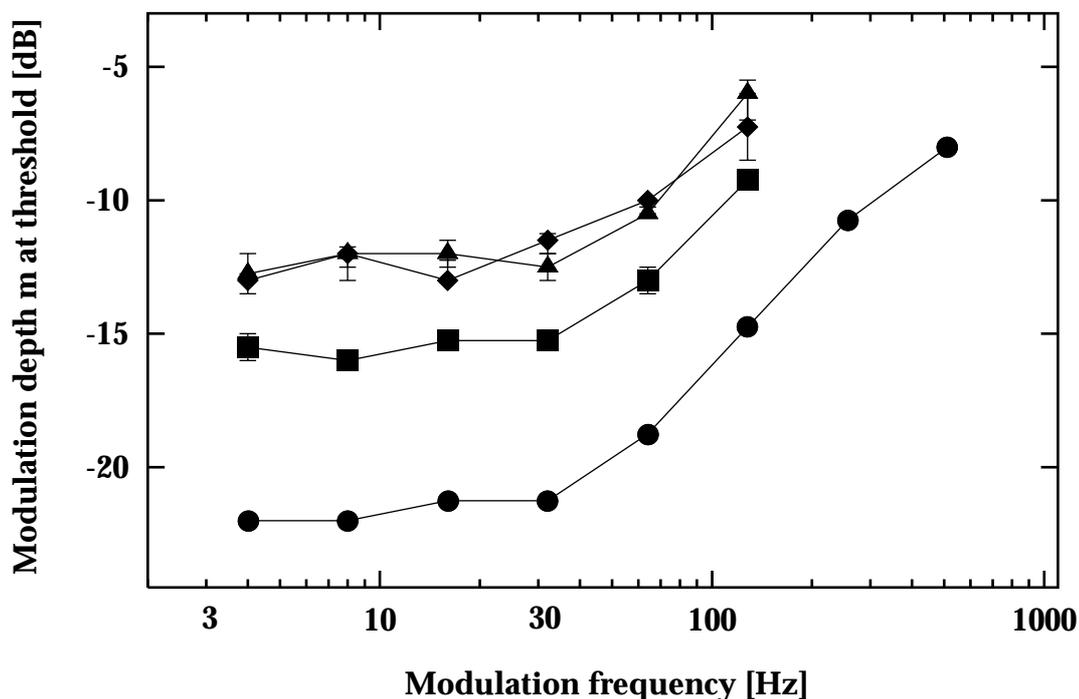


Figure 2.16: Simulations on the basis of Viemeister's model. Predicted modulation detection thresholds are shown for three different bandwidths of the noise carrier. Center frequency of the carrier: 5 kHz. Carrier bandwidth: ♦: 3 Hz, ▲: 31 Hz, ■: 314 Hz, ●: 6000 Hz.

Figure 2.16 shows simulated TMTF's for the noise carriers of 3, 31, 314 and 6000 Hz bandwidth on the basis of Viemeister's model. The stimuli were the same as in Sect. 2.5.1.1. All curves show a lowpass characteristic with a similar cut-off frequency. This characteristic reflects the influence of the low-pass filter stage. With decreasing carrier bandwidth the simulated TMTFs shift towards higher thresholds. This increase is a consequence of the spectral distribution of the intrinsic fluctuations in narrow-band noise (see Sect. 2.3). With decreasing carrier bandwidth, the variance of the unmodulated noise carrier increases. For the output of Viemeister's model, this increase will be seen for carrier bandwidths that are below the bandwidth of the predetection filter and above about the bandwidth of the lowpass filter.

This increase in threshold agrees with the experimental data for very low modulation frequencies. At higher modulation frequencies, however, the model predicts a totally different threshold pattern than that observed experimentally. While the pattern of the experimental data varies systematically with increasing carrier bandwidth, the model always predicts a lowpass characteristic in the threshold function independent of the carrier bandwidth.

Figure 2.17 shows model predictions of amplitude modulation detection thresholds of third-octave noise bands as a function of the center frequency. The stimuli were the same as in Sect. 2.5.1.2. The ordinate indicates modulation depth at threshold. Thresholds decrease monotonically with increasing center frequency.

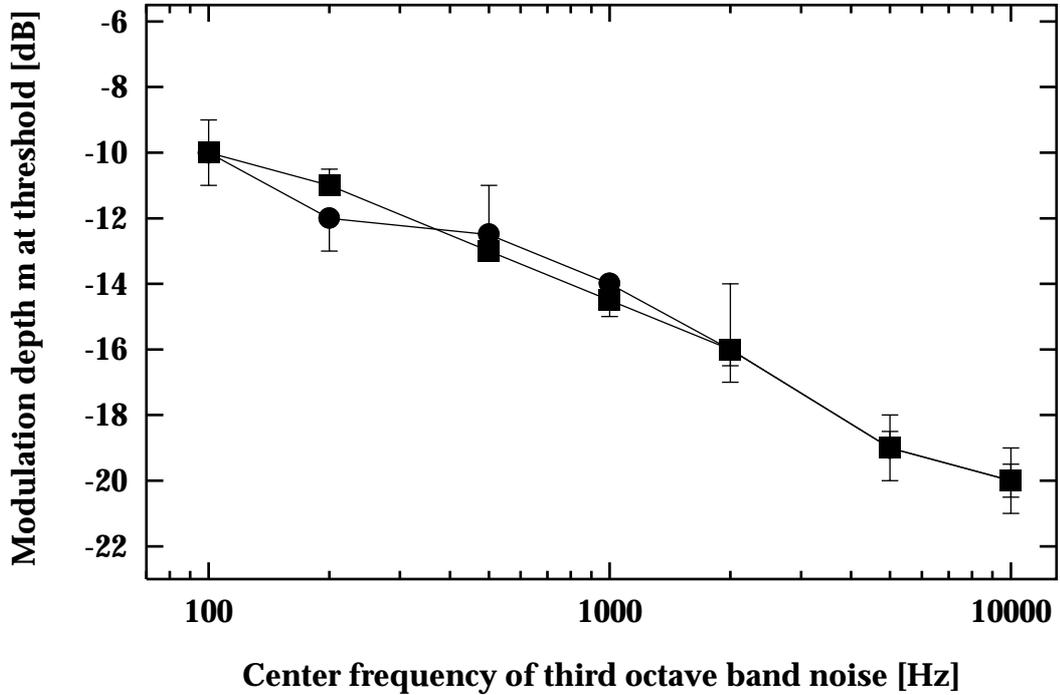


Figure 2.17: Simulated modulation detection thresholds of 25 Hz amplitude modulation as a function of the center frequency of the third octave band noise carrier. Viemeister model: (■), modulation filterbank model (●).

This is again caused by the decreasing variance of the unmodulated noise with increasing center frequency (and increasing linear bandwidth) at the output of the modulation-lowpass stage, in the same way as described above. The figure also shows the simulated data obtained with the modulation filterbank model (replot from Fig. 2.9). There is virtually no difference between the predictions of the two models. The predicted threshold functions account well for the data (see Sect. 2.5.1.2).

Finally, model predictions were calculated for the modulation masking experiment with the harmonic tone complex as the masker. The stimuli were the same as in Sect. 2.5.1.3. The model of Viemeister predicts some amount of masking (around 5 dB) which is the same for *all* test modulation frequencies. This frequency-independent amount of masking is caused by the specific model structure: Because there is only one modulation-frequency specific channel (the output of the low-pass filter) the effect of the masker modulation must be the same for all

test modulation frequencies. The experimentally observed increase in modulation masking with decreasing spectral distance between modulation masker and test modulation can thus be seen as strong support for a modulation-frequency specific analysis in the auditory system. The modulation-filter bank model, which is able to reproduce at least the trend in the data, is one possible realization for this analysis.

2.7 Discussion

The main goal of this study was to develop a model which describes the “effective” processing of envelope fluctuations in the auditory system. Experiments concerning modulation detection and modulation masking were performed which suggest that the auditory system realizes some kind of spectral decomposition of the temporal envelope of the signals. There seem to be channels in the auditory system which are tuned to modulation frequency, much like there are channels (critical bands or auditory filters) tuned to spectral frequency.

With regard to the experiments performed and the structure of the model that is inferred from this data, the following points have to be discussed: (a) The role of peripheral filtering for the processing of amplitude modulation, (b) the concept of a modulation filterbank in each critical band as opposed to a modulation lowpass filter, and (c) the envelope statistics of the different noise maskers employed and their influence on the thresholds obtained here.

With regard to point (a), it should be noted that the experiments in this study have been designed in such a way that effects of peripheral filtering can be neglected because the bandwidth of the modulated signals was always smaller than the bandwidth of the stimulated peripheral filter. However, it is often mentioned (e.g., Viemeister, 1979) that “off-frequency” listening may play a role in experiments on modulation detection of amplitude-modulated tones. The observer may increase the relative modulation depth by positioning his “internal observation filter” away from the carrier frequency in such a way as to better equate the amplitude of the carrier and one of the sidebands. The increase in relative modulation depth resulting from listening off frequency would improve performance. The magnitude of improvement would then depend on the attenuation characteristics of the peripheral filter. With increasing modulation frequency this improvement would generally increase. Within the present study, in most cases a center frequency of 5 kHz was applied and the highest test modulation frequency used was 150 Hz. Within this range of modulation frequencies (at least in the range 0-100 Hz), it is unlikely that off-frequency listening would be advantageous, and the subject most likely monitors the internal filter at the carrier frequency. This is supported by the observation that thresholds for a sinusoidal carrier *increase* beyond a modulation rate of 150 Hz (Fassel and Kohlrausch, 1995). Nevertheless, for higher modulation frequencies, when sideband cues become more and more available, a different detection strategy is probably used. In modulation-detection conditions with carrier bandwidths larger than a critical band, the influence of the peripheral filtering on the processing of modulation frequencies can no longer be neglected. Instead, an extension of the “single-channel” model is required that allows integration of signal information across frequency. Such an extension of the single-channel model to a “multi-channel” model, that is able to describe effects of spectral integration in amplitude-modulation detection and masking, is described in the next chapter.

With regard to point (b), it should be noted that the modulation filterbank concept proposed here is different both from the previous version of the current model (cf. Dau et al., 1996a,b) and from models proposed in the literature (e.g., Viemeister, 1979). Both of these models employed some kind of a modulation lowpass filter. The previous version of the current model was developed to describe simultaneous and forward masking data and included a lowpass filter with a cut-off frequency of 8 Hz. Such lowpass filtering, however, would fail to describe experiments concerning modulation masking and would also fail to describe basic experiments concerning modulation detection with narrowband carriers at high center frequencies (cf. Sec. 2.5.1.1). The present model allows the prediction of modulation data and, at the same time, preserves the capabilities of the earlier model for describing simultaneous and nonsimultaneous masking data. This is because the linear modulation lowpass filtering (with a cut-off frequency at very slow modulations) is retained in the current model and is combined with the analysis of faster modulations by a modulation filterbank (see also the simulations with a slightly different modulation-filterbank implementation in Münkner and Püschel, 1993). The idea behind the modulation lowpass-filter approach described in the literature (cf., Viemeister, 1979) is that a “minimum integration time” is typically derived from the cut-off frequency of the lowpass characteristic in the threshold function as a parameter that describes the auditory system’s temporal resolution (for a review, see Viemeister and Plack, 1993). Such a model is capable of predicting a variety of different experiments, for example, the TMTF for broadband noise carriers (cf. Fig. 2.7) and modulation thresholds in third-octave band noise at different center frequencies (cf. Fig. 2.9). However, a model employing only a lowpass filter fails to describe the modulation detection data for a narrowband-noise carrier. Furthermore, such a model fails to describe the masking data using the tone-complex modulation masker. Hence, the model proposed here considerably expands the class of experiments that can be modeled correctly while still maintaining some of the properties and predictions of the model proposed by Viemeister (1979).

With regard to point (c) it should be noted that there are only a few studies which have attempted to involve the *inherent* statistical properties of the noise carriers (e.g., Zwicker, 1953; Maiwald, 1967a,b; Fleischer, 1981, 1982b). For example, Fleischer (1981, 1982a,b) investigated TMTFs using narrowband noise as the carrier. He developed a model for describing the interaction between inherent fluctuations within a noise carrier and the detectability of added modulation. The “modulation spectrum” was weighted by a certain factor which essentially represented a low-pass characteristic. For modulation frequencies lower than half the bandwidth of the noise carrier, this model yields good agreement with experimental results for modulation detection and modulation difference limens (Fleischer, 1981, 1982a,b). For modulation frequencies larger than half the bandwidth of the noise carrier, this model would always predict a lowpass characteristic in the threshold function without regard to the applied carrier bandwidth - in the same

way as shown in Sect. 2.6 for the Viemeister model. Therefore, to account for the data, Fleischer extended the model by assuming a “cross-talk” between the inherent fluctuations of the noise and the added modulation. He postulated a decay at a rate of 16 dB per decade of the modulation frequency to account for the highpass characteristic in the data. In order to find a description for the inherent modulation of the noise, Fleischer (1981) regarded narrowband noise with a bandwidth Δf as a pure tone which was amplitude modulated by a continuum of equal-amplitude modulation frequencies between zero and half the bandwidth of the noise. But this assumption is not correct. It would imply that the modulation spectrum of noise has a flat shape in contrast to the triangular shape, as shown by Lawson and Uhlenbeck (1950) (see Sect. 2.3).

Even though the exact shape of the modulation spectrum assumed by Fleischer (1981) was not correct and the subsequent explanation of the data was based on a different concept than the one described here, Fleischer’s concept of “cross talk” between inherent envelope fluctuations of the carrier noise and the test modulation is compatible with the bandpass analysis proposed here. Within the modulation filterbank model, the lowpass characteristic of the threshold function for conditions with $f_{mod} < \frac{\Delta f}{2}$ does not result from a specific weighting function used to model the “sluggishness” of the auditory system. Instead, it is a consequence of the intrinsic envelope fluctuations of noise bands and their spectral distribution on the one hand, and of the logarithmic scaling of the postulated modulation filters on the other hand. The current model can also account for the data using very narrowband stimuli as the carrier, describing a highpass or bandpass characteristic in the threshold function.

2.8 Conclusions

(1) – The experiments on modulation detection and modulation masking described here agree well with experiments from the literature. They all provide evidence for an analysis of envelopes in terms of a separation into different modulation frequencies.

(2) – The model of the “effective” signal processing in the auditory system proposed here is capable of quantitatively modeling most aspects of the experiments described. It employs a modulation filterbank for envelope analysis that exhibits a constant absolute bandwidth for low frequencies and a constant relative bandwidth for modulation frequencies above 10 Hz.

(3) – While the model proposed here coincides with some structures and predictions of the modulation lowpass model by Viemeister (1979) and an earlier version of the current model (Dau et al., 1996a,b), it still holds for experiments where the modulation lowpass approach clearly fails (e.g., modulation detection for narrowband-noise carriers).

Chapter 3

Spectral and temporal integration in amplitude modulation detection¹

Abstract

A multi-channel model, describing the effects of spectral and temporal integration in amplitude-modulation detection for a stochastic noise carrier, is proposed and validated. The model is based on the modulation-filterbank concept which was established in the accompanying paper by Dau et al. (1996c) for modulation perception in narrowband conditions (single-channel model). To integrate information across frequency, the detection process of the model optimally weights and linearly combines the channel outputs, assuming independent observations at the outputs. To integrate information across time, a “multiple-look” strategy is realized within the detection stage of the model. Both data from the literature and new data are used to validate the model.

The model accounts for the findings by Eddins [J. Acoust. Soc. Am. 93, 470–479 (1993)] that the “time constants” associated with the temporal modulation transfer functions (TMTF) derived for narrowband stimuli do not vary with carrier frequency region and that the time constants associated with the TMTFs decrease monotonically with increasing stimulus bandwidth. The model is able to predict masking patterns in the modulation-frequency domain, as observed experimentally by Houtgast [J. Acoust. Soc. Am. 85, 1676–1680 (1989)]. The model also accounts for the finding by Sheft and Yost [J. Acoust. Soc. Am. 88, 796–805 (1990)] that the long “effective” integration time constants derived from the data are two orders of magnitude larger than the time constants derived

¹Modified version of the paper “Modeling auditory processing of amplitude modulation: I. Spectral and temporal integration”, written together with Birger Kollmeier and Armin Kohlrausch, submitted to J. Acoust. Soc. Am.

from the cut-off frequency of the TMTF. Finally, the temporal-integration properties of the model allow the prediction of data in a specific temporal summation paradigm used earlier by Wakefield and Viemeister [(J. Acoust. Soc. Am. 90, 858–865 (1991))].

The combination of the modulation filterbank concept and the optimal decision algorithm proposed here appears to present a powerful strategy for describing modulation detection phenomena in narrowband and broadband conditions.

3.1 Introduction

One of the most interesting and fundamental questions in psychophysical research is how the auditory system “trades” spectral and temporal resolution. One problem in this field is the question of how peripheral filtering affects the ability to detect modulation. It is often postulated that peripheral filtering does not limit the ability to detect modulation and that in “temporal resolution” tasks such as modulation detection the observer broadens his “effective” bandwidth (e.g., Viemeister, 1979). With such an assumption, the experimental data can be simulated by a model using a “predetection” filter as broad as several critical bandwidths to account for some kind of peripheral filtering (Viemeister, 1979). If, in contrast, peripheral filtering in terms of critical bands has any influence on modulation detection, the question arises how the applied modulation analysis depends on center frequency. In this vein, some authors have postulated that the TMTF of a broadband noise is actually determined by the information in the *highest* critical band excited by the stimulus (e.g., Maiwald, 1967a,b; Van Zanten, 1980). Eddins (1993) examined spectral integration in amplitude-modulation detection, independently varying stimulus bandwidth and spectral region. He found that these data could not be described by models from the literature (e.g., Viemeister, 1979).

In the previous chapter, experiments on amplitude modulation detection were described using narrowband noise as the carrier at a high center frequency (5 kHz). By using these conditions, effects of any peripheral or “predetection” filtering were minimized and, in addition, the spectral region that was being used to detect the modulation was restricted to one critical band. A model of the “effective” signal processing in the auditory system, including a modulation filterbank, was derived to analyze the temporal envelope of the stimuli. This model will be called the “modulation filterbank model” throughout this chapter.

To get more insight into the processing of modulation, several experiments are described here that investigate the effects of spectral integration in amplitude-modulation detection, examining in particular the transition between narrowband and broadband conditions. To test the capabilities of the modulation filterbank model in conditions of spectral integration, some critical experiments are performed and model predictions are compared with data from the literature (Eddins, 1993; Sheft and Yost, 1990) and the new experimental data. To compare results from experiments and simulations with as close a similarity in the experimental details as possible, the experiments by Eddins (1993) and those by Sheft and Yost (1990) were replicated with a slightly different threshold estimation procedure.

Another aspect of modulation perception is the phenomenon of temporal integration in amplitude modulation detection. Temporal integration or temporal summation refers to the well-known fact that over a range of durations there is a decrease in threshold with increasing signal duration. Several models have been

suggested in the literature to describe this phenomenon. However, differences in the modeling strategies occur that reflect the “resolution-integration conflict” (de Boer, 1985). In the case of temporal integration in modulation detection, Sheft and Yost (1990) found that time constants associated with the slope of the temporal integration function per unit duration are much larger than those indicated by the “resolution data”. The present study examines the ability of the modulation filterbank model to account for these long “effective” time constants found in the data.

As before, both the own experimental data from “critical” experiments as well as data from the literature were compared with model predictions. It should be emphasized at this point that the same parameter set was used for all model predictions. These parameters were partly derived from earlier experiments and partly derived from a subset of the data described here.

3.2 Method

3.2.1 Procedure and Subjects

Modulation thresholds were measured and simulated using an adaptive 3-Interval Forced-choice (3IFC) procedure. The unmodulated carrier was presented in three consecutive intervals, separated by silent intervals of 300 ms. In one randomly chosen interval the carrier was sinusoidally amplitude modulated. The subject's task was to specify the interval containing the test modulation. During a threshold run, the modulation depth in dB ($20 \log m$), was adjusted by a 2-down 1-up algorithm (Levitt, 1971) which provides an estimate of the modulation depth necessary for 70.7% correct responses. The step size was 4 dB at the start of a run and was divided by 2 after every two reversals of the signal level until the step size reached a minimum of 1 dB, at which time it was fixed. Using this 1-dB step size, ten reversals were obtained and the median value of the signal levels at these 10 reversals was used as the threshold estimate. The subjects received visual feedback during the measurements. The procedure was repeated four times for each signal configuration and subject. All figures show the median and interquartile ranges based on four single measurements. Five subjects participated in this study. They had some prior experience in psychoacoustic measurements and had normal hearing. They were between 23 and 29 years old.

3.2.2 Apparatus and stimuli

All acoustic stimuli were digitally generated at a sampling frequency of 30 kHz. The stimuli were transformed to analog signals with a two-channel 16-bit D/A converter, attenuated, lowpass filtered at 10 kHz and diotically presented via headphones (HDA 200) in a soundproof booth. Signal generation and presentation were controlled by a SUN-Workstation using a signal-processing software package developed at the Drittes Physikalisches Institut in Göttingen.

Several experiments on modulation detection were performed. The subject's task was to detect a sinusoidal test modulation of a Gaussian noise carrier. Unless explicitly stated, a carrier level of 65 dB SPL was used. During the experiments with a noise carrier, an independent sample of the noise was presented in each interval. The carrier noise stimuli were digitally filtered in the frequency domain by Fourier transforming the whole noise waveform, setting the respective frequency samples to zero, and transforming the signal back into the time domain. The rms level of the modulated signal was always equated to the rms level employed in the unmodulated trials.

3.3 Multi-channel model

All simulations discussed in this chapter were performed on the basis of the modulation filterbank model that was initially developed as a single-channel model as described in the previous chapter. Figure 3.1 shows the structure of the model which was extended to a multi-channel model in such a way that the modulation analysis is performed in parallel on the output of each stimulated peripheral filter. This may be motivated in part by results from physiological studies of

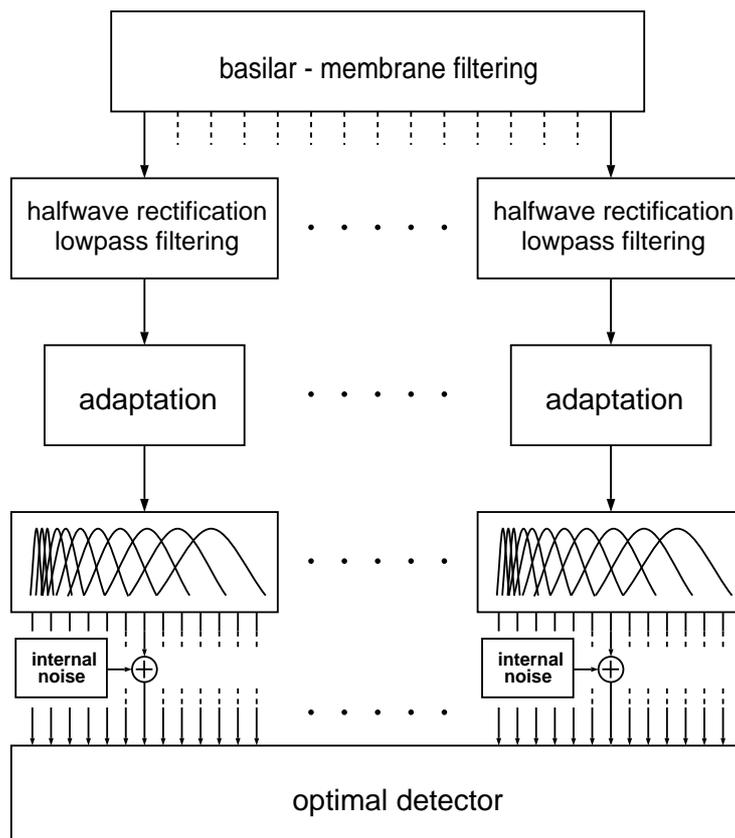


Figure 3.1: *Block diagram of the psychoacoustical model for describing modulation detection data with an optimal detector as decision device. The signals are preprocessed, subjected to adaptation, filtered by a modulation filterbank and finally added to internal noise; this processing transforms the signals into their internal representations.*

the representation of amplitude modulation within the Inferior Colliculus (IC) of the cat, where it was found that modulation frequencies are represented in a systematic fashion orthogonal to the tonotopic organization of the IC (Schreiner and Langner, 1988).

The model contains the same stages of signal processing along the auditory

pathway as proposed in the single-channel model. The main features are described briefly here. For further details see Dau et al. (1996a,c). To simulate the peripheral bandpass characteristic of the basilar membrane the gammatone filterbank of Patterson et al. (1987) is used. The filters overlap at the -3-dB points of their transfer functions. The summation of all filter transfer functions would result in a flat transfer characteristic across frequency. In conditions with broadband noise as the carrier, the filtered waveforms of the noise at the outputs of different peripheral filters may be considered as statistically independent of each other. The stimulus at the output of each peripheral filter is half-wave rectified and lowpass filtered at 1 kHz. This stage roughly simulates the transformation of the mechanical oscillations of the basilar membrane into receptor potentials in the inner hair cells. The lowpass filtering essentially preserves the envelope of the signal for high center frequencies. Effects of adaptation are simulated by nonlinear adaptation circuits (Püschel, 1988; Dau et al., 1996a,b). In a stationary condition, the transformation has a compressive characteristic which is close to logarithmic. Fast fluctuations of the input are transformed more linearly. The adapted signal is further analyzed by a linear modulation filterbank. The parameters of the modulation filters have been determined in the previous study by modeling modulation data for subcritical stimulus bandwidths. The filterbank has two domains with different scaling. In the range of center frequencies 0 – 10 Hz a constant absolute bandwidth of 5 Hz is used. Above 10 Hz up to 1000 Hz a logarithmic scaling with a constant Q-value of 2 is used. For modulation filters tuned to modulation frequencies larger than 10 Hz, only the Hilbert envelope of the output is further analyzed as motivated in the previous study. It is assumed within the present study that these parameters do not vary across frequency, i.e., the same modulation filterbank is applied to analyze the signal's envelope at the output of each critical band. Limitations of resolution are simulated by adding internal noise with a constant variance to each modulation filter output. The decision device is realized as optimal detector in the same way as described in Dau et al.(1996a,b). In the decision process a stored temporal representation of the signal to be detected (template) is compared with the actual pattern. The comparison amounts to calculating the cross correlation between the two temporal patterns and is comparable to a “matched filtering” process. Assuming independent “observations” across frequency, the model combines all filter outputs linearly and an optimal weighting of the channels is used. The optimality of the detection process refers to the best possible theoretical performance in detecting the signal under specific conditions, including the added internal noise. Using such a model, a prediction of the average subject's performance is possible on a trial-by-trial basis. Simulations of the experimental runs can thus be performed by implementing on the computer the same threshold estimating procedures within the model as those being used in the experiments with human observers.

3.4 Results from measurements and simulations

3.4.1 Modulation analysis within and beyond one critical band

In the first experiment a noise carrier centered at 5 kHz was sinusoidally modulated with a modulation rate of 5 Hz. Modulation depth at threshold was measured as a function of the carrier bandwidth, which had one of the following values: 10, 25, 50, 100, 250, 500, 1000, 1500, 2500, 5000 or 10000 Hz. The durations of the carrier and the modulation were 500 ms including 50-ms Hanning-shaped rise and decay ramps. The bandwidths cover the range from less than the critical bandwidth at 5 kHz to much greater than the critical bandwidth. The left panel of Fig. 3.2 shows the experimental results of three subjects (open symbols). The ordinate indicates modulation depth at threshold and the abscissa shows the carrier bandwidth. Thresholds decrease monotonically with increasing

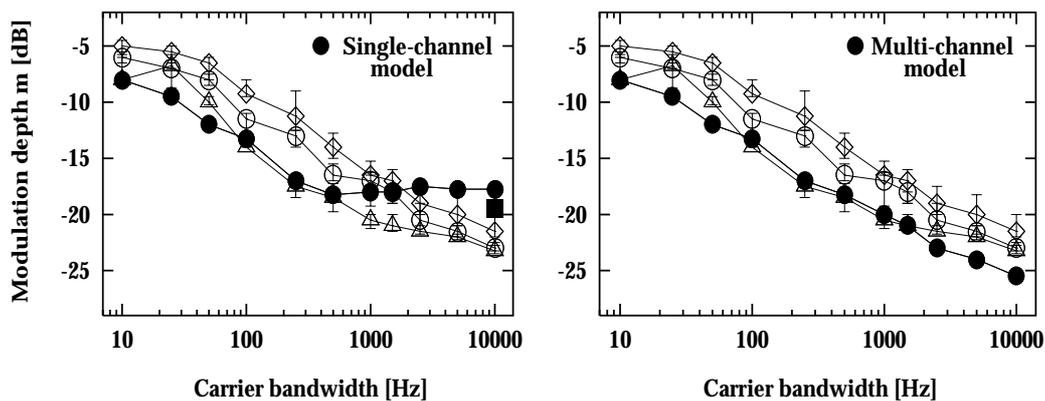


Figure 3.2: Modulation detection thresholds of a 5-Hz modulation as a function of the carrier bandwidth. Open symbols indicate measured data of three subjects (same in the left and the right panel). Simulated thresholds with the single-channel model are represented in the left panel by the filled circles. Simulated thresholds with the multi-channel model are shown in the right panel (filled circles). The filled box represents the simulated threshold for $\Delta f = 10$ kHz, where the modulation analysis was performed using the highest critical band excited by the stimulus. Center frequency: 5 kHz; Carrier and modulation duration: 500 ms. Level: 65 dB SPL. Subjects: JV (\square), TD (\circ), PD (\diamond), optimal detector (\bullet , \blacksquare).

bandwidth. This is due to the decreasing spectral energy density of the inherent fluctuations with increasing bandwidth of the carrier, when the total energy of the modulated signal is constant. As a consequence, less energy of the random

envelope fluctuations of the carrier leaks into the transfer range of the modulation filters near the test modulation frequency. In the model, this leads to a decreasing threshold with increasing carrier bandwidth. Simulated thresholds obtained with the single-channel model as described in the previous chapter are plotted in the left panel of Fig. 3.2 (filled symbols). In this single-channel simulation, the modulation analysis was carried out using the peripheral filter tuned to 5 kHz. In conditions with carrier bandwidths $\Delta f < 1000$ Hz, simulated and measured data are in good agreement. For bandwidths $\Delta f \geq 1000$ Hz, however, systematic differences between simulations and experimental data occur, which increase with increasing bandwidth. For the bandwidth $\Delta f = 10$ kHz, for example, the difference between experiment and simulation amounts to more than 5 dB. Of course, this discrepancy is expected, since the simulations were carried out using only the information about the signal modulation available at the output of one peripheral filter. The filled box in the left panel of Fig. 3.2 represents the simulated threshold for a carrier bandwidth of 10 kHz, where the modulation analysis was performed using the highest critical band excited by the stimulus. Even in this “optimized” single-channel simulation, there remains a discrepancy of nearly 5 dB between the simulated and the measured threshold. This does not support the hypothesis by Maiwald (1967) and Van Zanten (1980) that sufficient information about the signal modulation is available in the highest excited frequency region and that the detection strategy of the subject is to monitor the “internal” filters in this high-frequency region.

The right panel of Fig. 3.2 shows the simulated data from the multi-channel filterbank model together with the experimental data already shown in the left panel. Using information across frequency, the model accounts for the data over the whole range of carrier bandwidths.

3.4.2 Effects of bandwidth and frequency region

In the next experiment, effects of absolute bandwidth and frequency region on spectral integration in modulation detection were studied using narrowband noise as the carrier. As in the study of Eddins (1993), modulation thresholds were measured as a function of modulation frequency for the following conditions: The stimulus bandwidth was either 200, 400, 800, or 1600 Hz and the frequency region was varied by adjusting the high-frequency cutoff of the noise to be either 600, 2200, or 4400 Hz. The purpose was to determine temporal modulation transfer functions (TMTF’s) by *independently* varying stimulus bandwidth and stimulus frequency to determine the influence of these parameters on modulation detection. As in Eddins (1993), the modulated stimuli were generated by bandpass filtering *after* amplitude modulation of wideband noise to avoid the possibility that the detection of modulation would be based on spectral cues in the signal interval rather than changes in the temporal waveform. In this way, the bandwidth of the narrowband stimuli was the same in the presence or absence of modulation.

In addition, the stimuli were adjusted to have equal energy in each interval of the forced-choice trial to prevent detection of modulation based on overall level rather than on the presence or absence of modulation.

Figures 3.3 and 3.4 show the experimental data of two subjects for several stimulus conditions. Modulation depth m at threshold is plotted as a function of modulation frequency. The transfer functions reflect a lowpass characteristic that is similar to data from previous studies (Rodenburg, 1972, 1977; Viemeister, 1977, 1979; Formby and Muir, 1988). The data are in very good agreement with the results of Eddins in spite of slight differences in the threshold estimation procedure between both studies. Increasing stimulus bandwidth results in a corresponding increase in sensitivity to modulation. This systematic increase is seen for each of the three cutoff frequencies tested. Eddins fitted his data -

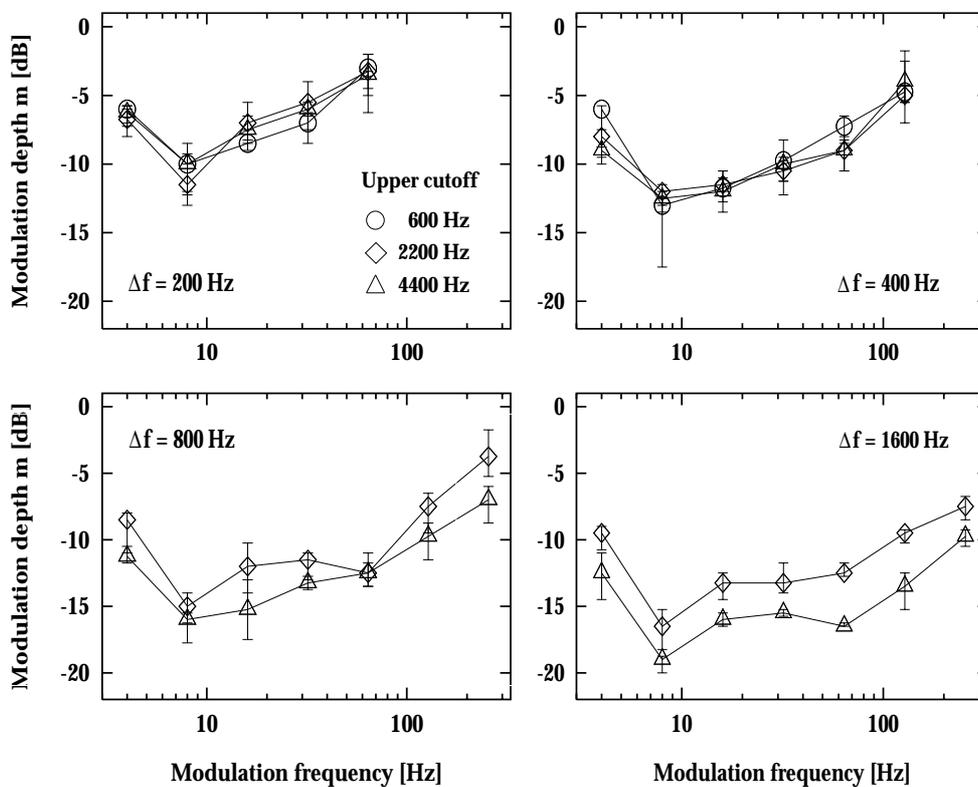


Figure 3.3: Modulation detection thresholds for one subject (BG) as a function of modulation frequency with upper cutoff frequency as the parameter. Each panel represents a different bandwidth condition: upper left: 200 Hz, upper right: 400 Hz, lower left: 800 Hz, lower right: 1600 Hz. The upper cutoff frequency is either 4400 Hz (\triangle), 2200 Hz (\diamond) or 600 Hz (\circ).

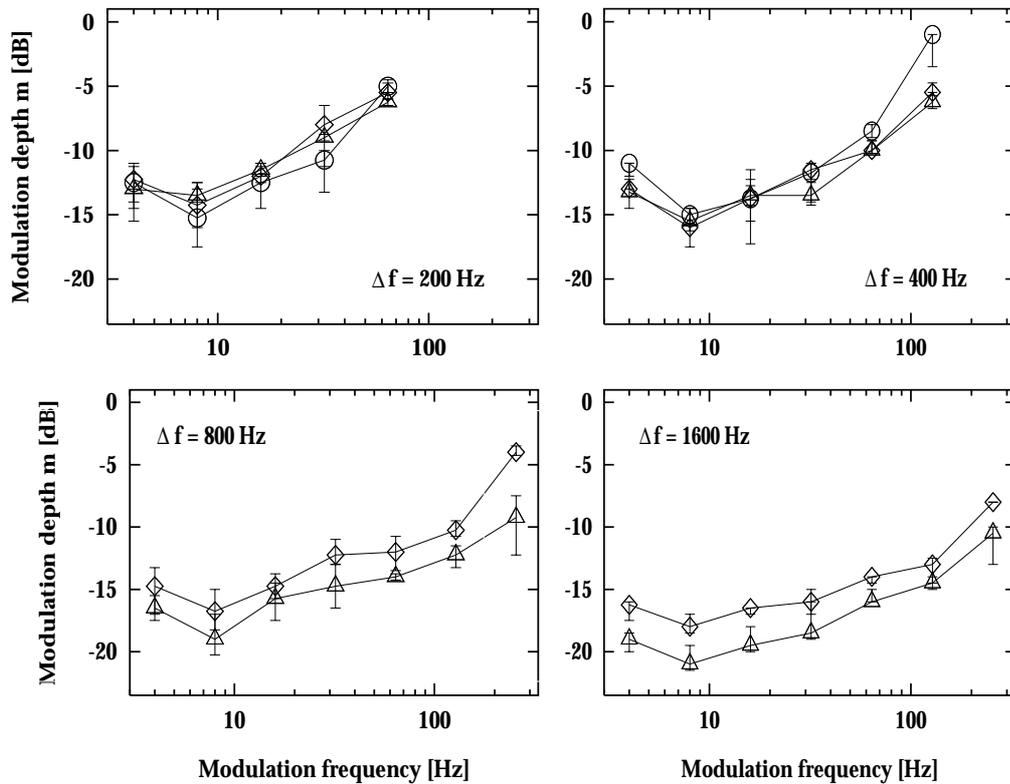


Figure 3.4: Same as Fig. 3.3 for subject TD.

for each subject and condition - with the transfer function of a simple lowpass filter and derived the time constant associated with the -3-dB cut-off frequency of the specific transfer function. Two important interpretations emerged from his analysis: First, the time constants associated with the TMTFs do not vary with changing frequency region. Second, the time constants associated with TMTFs decrease monotonically with increasing stimulus bandwidth. Eddins (1993) concluded from the data that temporal acuity is independent of frequency region, assuming that temporal acuity and derived time constants from the data are directly related.

Figure 3.5 shows the corresponding model predictions obtained with the modulation filterbank model. Modulation depth at threshold is plotted as a function of modulation frequency. The shape of the simulated threshold patterns agrees well with the corresponding data. That is, the *form* of the simulated TMTF does not depend on the frequency region of the stimuli but depends on the stimulus bandwidth. In one - “extreme” - condition, for $\Delta f = 400$ Hz at the upper cut-off frequency of 600 Hz, the slope of the simulated TMTF between the penultimate and the last modulation frequency is steeper than in case of the two other cut-

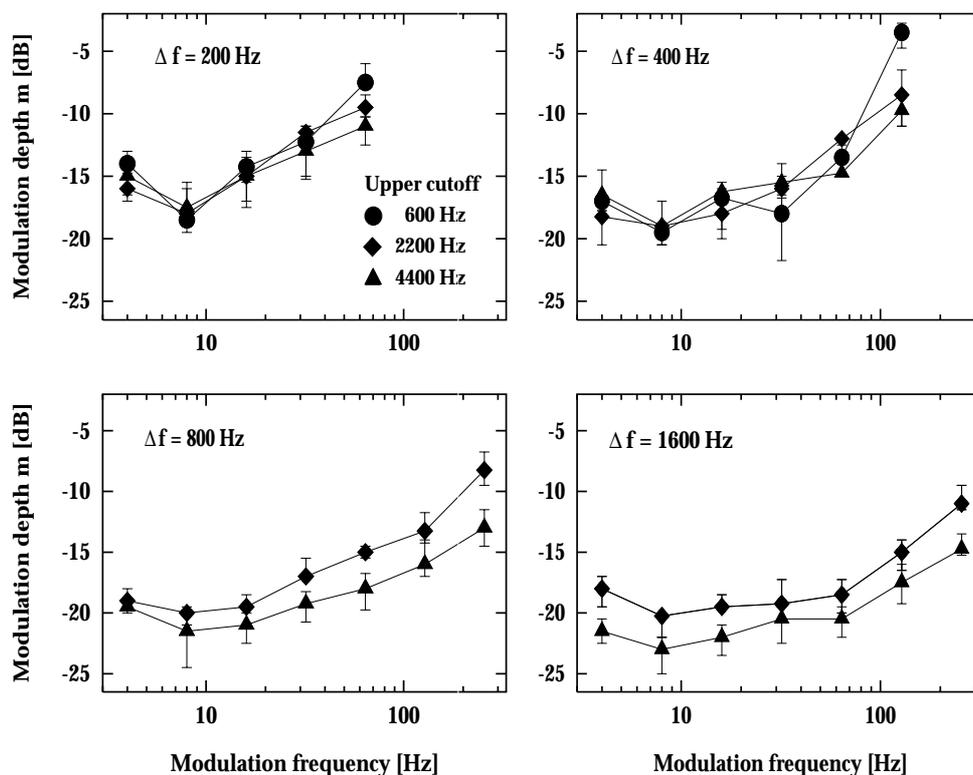


Figure 3.5: Simulated modulation detection thresholds as a function of modulation frequency for the same conditions as in Figs. 3.3 and 3.4. Thresholds are indicated by filled symbols.

offs at this bandwidth, and is steeper than the corresponding experimental data in this condition. Note, however, that one of the subjects (TD) shows an increase of 6 dB between the penultimate and the last modulation frequency in this condition.

Within the model the “low-frequency” and the “high-frequency” conditions require quite different decision strategies to determine the detection threshold. Let us consider, for example, the bandwidth of $\Delta f = 400$ Hz and the upper cutoff frequencies of 600 Hz and 4400 Hz, representing the low-frequency and the high-frequency conditions, respectively. In the low-frequency condition, the modulation analysis is performed in parallel in several peripheral filters. The transfer functions of the adjacent gammatone filters are plotted in Fig. 3.6. In the same figure, the transfer function of the gammatone filter tuned to the high-frequency region is also plotted. Because of the poor spectral resolution of the auditory system at high frequencies, the modulation analysis is carried out in only one peripheral channel. The results of the simulations in the modulation detection

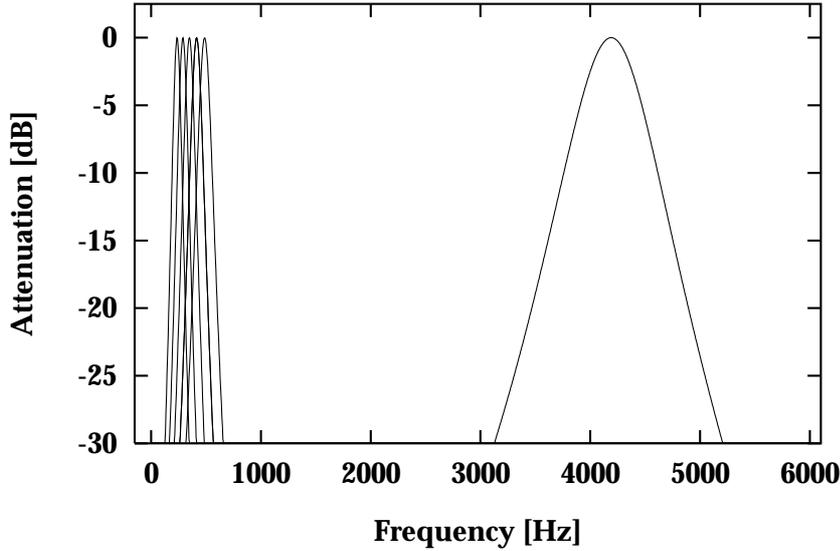


Figure 3.6: Transfer functions of the gammatone filters of Patterson et al. (1987) in the low-frequency region (200 – 600 Hz) and the high-frequency region (4000 – 4400 Hz) plotted on a linear frequency scale.

task are the same for both conditions, indicating that, apparently, the poor temporal resolution in the low-frequency condition is somehow compensated for by the greater number of independent “observations” across frequency compared to the high-frequency condition.

According to signal detection theory (SDT) (e.g., Green and Swets, 1966), it is equivalent to detect the modulation of a single noise at the output of one wide peripheral filter or to combine detection information across an array of bandpass-filtered uncorrelated noises. This is true in the conditions with stochastic noise carriers and a coherent deterministic amplitude modulation. This implies *independent* individual observations across frequency. When n observations are combined, the value of the detectability index d' , denoted d'_n , is equal to the square root of the sum of the squares of the individual values of d'_i .²

$$d'_n = \sqrt{\sum_i^n (d'_i)^2} \quad (3.1)$$

In deriving this formula, the decision mechanism used in the model realizes an optimal linear combination of all filter outputs at the level of the internal repre-

²Assuming $d' = \sqrt{\frac{2E}{N}}$ from SDT where S denotes signal energy and N denotes noise energy, and with $E_i = \frac{E}{n^2}$ for coherent signals and $N_i = \frac{1}{n}N$ for uncorrelated noises, it is:
 $\sqrt{\sum_i^n (d'_i)^2} = \sqrt{\sum_i^n \frac{2E_i}{N_i}} = \sqrt{\sum_i^n \frac{2E}{n^2 N_i}} = \sqrt{\sum_i^n \frac{2E}{N} \frac{1}{n}} = \sqrt{\sum_i^n \frac{2E}{N}} = d'_n$.

sentation of the stimuli.³

A factor limiting the validity of the above formula is the absolute threshold for detecting modulation of a certain rate. Such a threshold was introduced within the present model by the addition of *internal* noise to the output of each modulation filter. For a sinusoidal carrier this internal noise results in a threshold of detectability for a low-frequency amplitude modulation of about -27 dB. If, in the case of a noise carrier, the imposed modulation is strongly attenuated within one or more peripheral filters, the detectability of the modulation is not determined solely by the inherent statistics of the noise carrier but is also determined by the internal noise. As a consequence, in such a condition, some of the (low-frequency) peripheral filters do not contribute to the information about the signal modulation leading to an increased detection threshold. Therefore, in an “extreme” condition as described above, namely in the case of the highest imposed modulation frequency in the lowest frequency region, the width of the peripheral filter should have an influence on modulation detection.

Apart from such specific conditions, however, there is a very good agreement between experiment and model achieved with the traditionally assumed width of the auditory filters.

3.4.3 Further experiments and analytical considerations

From the previous section it can be concluded that, for broadband noise carriers, there is little influence of peripheral filtering on modulation detection. Thus, the changes in TMTF shape with increasing carrier bandwidth must be caused by the properties of “higher” processing stages in the auditory pathway. The results of the experiments from the first part of the previous chapter dealing with stimulus bandwidths smaller than a critical band could be explained as masking effects in the modulation-frequency domain. A quantitative description of the data was possible when the modulation-frequency selectivity was modeled by a modulation filterbank. Depending on the actual carrier bandwidth, the simulations led to a highpass or lowpass characteristic of the threshold pattern, in very good agreement with the data.

To understand better the transition between narrowband and broadband conditions an additional experiment was performed, in which carrier bandwidths were varied on a finer scale than in previous studies and only a few modulation frequencies were imposed. Figure 3.7 shows experimental data for three subjects, whose task was to detect a modulation of 5, 25 or 100 Hz that was imposed on a noise carrier with a level of 65 dB SPL with a fixed upper-cutoff frequency of 6 kHz. The parameter was the carrier bandwidth which was either 1, 5, 10, 25, 50, 100, 250, 500, 1000, 4000 or 6000 Hz. In all conditions of this experiment, the

³This is equivalent to assuming that the individual observations are combined via the likelihood ratio (see Appendix A; Green and Swets, 1966).

modulated stimuli were generated by modulating *after* bandpass filtering. The stimuli were shaped with 50 ms Hanning ramps.

The figure shows modulation depth at threshold as a function of the carrier bandwidth. The parameter in this plot is the imposed modulation frequency ($f_{mod} = 5, 25, 100$ Hz). The figure shows three “bandpass-shaped” curves. For each curve, the maximum occurs at a bandwidth equal to 2–5 times the imposed modulation frequency.

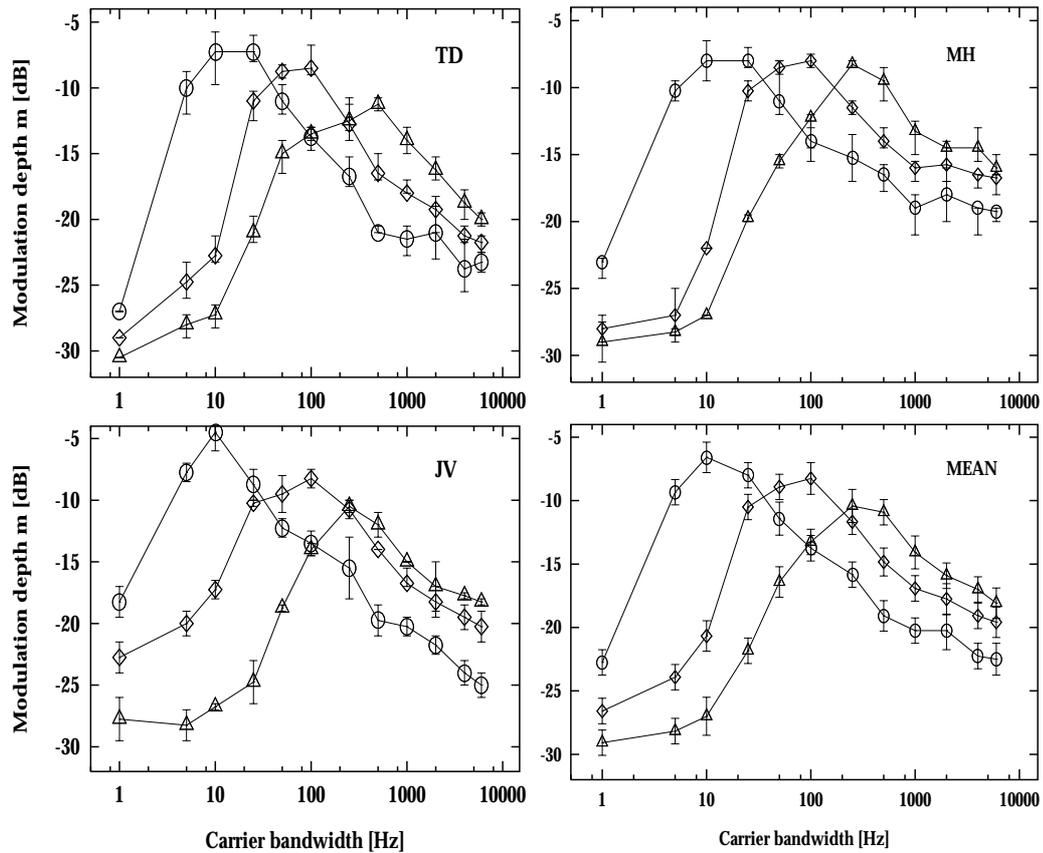


Figure 3.7: Modulation detection thresholds as a function of the carrier bandwidth. Parameter is the signal modulation frequency: $f_{mod} = 5$ Hz (\circ), $f_{mod} = 25$ Hz (\diamond), $f_{mod} = 100$ Hz (\triangle); Subjects: TD (upper left), MH (upper right), JV (lower left), mean (lower right).

In the following, the experimental data are compared analytic calculations. They are based on analytical properties of the statistics and spectral distribution of power in the modulation spectra of noise signals, as derived in the previous chapter. For the calculation, it is first assumed that peripheral filtering does not play a role in modulation detection in the stimulus conditions considered

here. Second, only the modulation filter tuned to the current signal modulation frequency is considered, assuming that the observer selects the modulation filter with the greatest signal-to-noise ratio at the output.

Figure 3.8 shows the transfer functions of the modulation filters that are tuned to the modulation frequencies 5, 25 and 100 Hz, respectively, as they are used in the model. The scaling of the modulation filters was chosen according to the results from the previous study. The functions are given on a log-log plot.

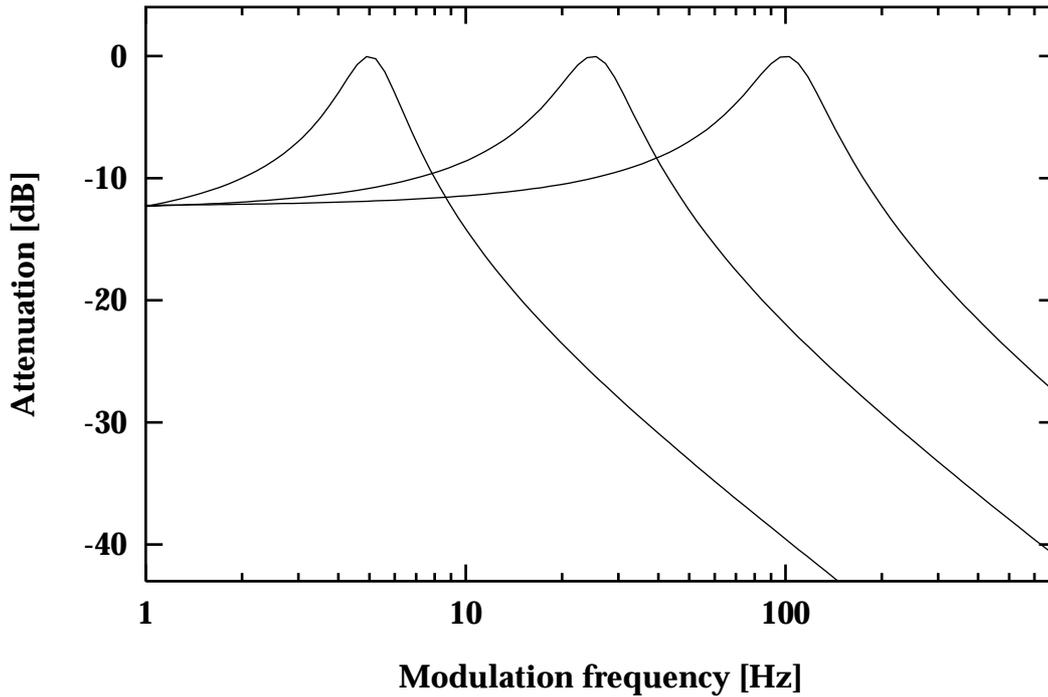


Figure 3.8: *Transfer characteristic of the modulation filters tuned to 5, 25 and 100 Hz, respectively. The parameters of the filters were chosen according to the previous chapter.*

To calculate the spectral power density of a specific noise within the transfer range of a specific modulation filter with modulation center frequency CF , the modulation power spectrum of the carrier, N , was multiplied by the squared amplitude transfer function of the modulation filter (from Fig. 3.8), denoted as $W_{CF} = W_{CF}(f_{mod})$. The modulation power spectra of the noise carriers were calculated according to the analytical calculation by Lawson and Uhlenbeck (1950) (see Sect. 4 of the previous chapter). They have approximately a triangle-like shape according to the formula:

$$N_{\Delta f, \rho}(f_{mod}) \approx \pi \Delta f \rho \delta(f_{mod}) + \frac{\pi \rho}{4 \Delta f} (\Delta f - f_{mod}), \quad (3.2)$$

where Δf is the carrier bandwidth, ρ is the spectral power density and f_{mod} indicates modulation frequency.

Figure 3.9 shows, as an example, the spectral power density as a function of modulation frequency considering only the modulation filter tuned to 25 Hz. The parameter in this figure is the noise bandwidth Δf . To determine the total

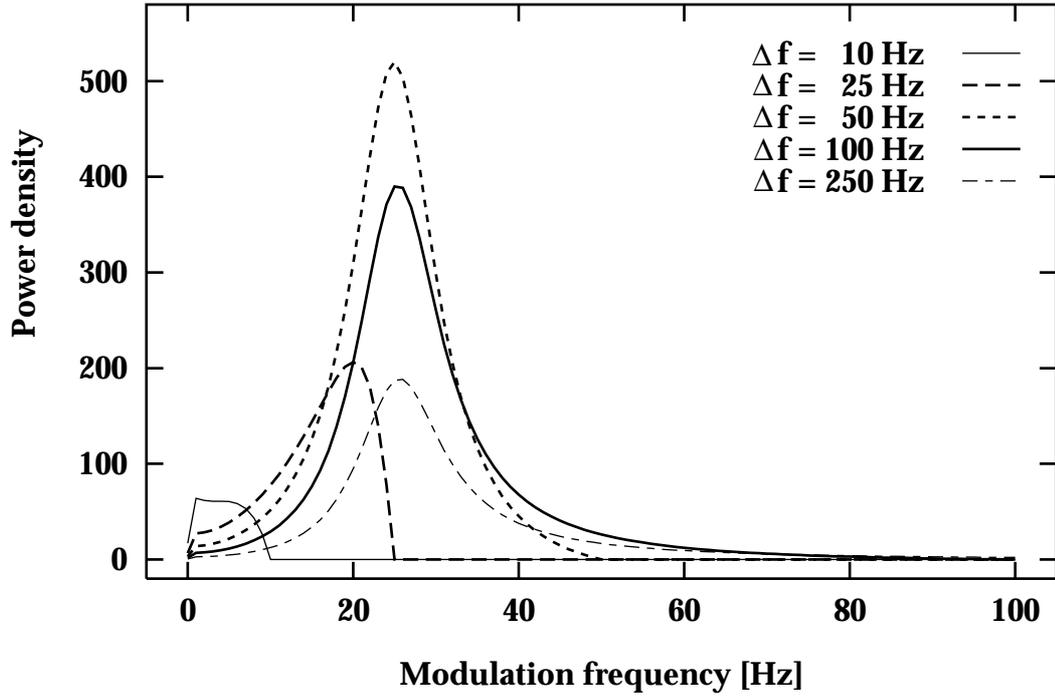


Figure 3.9: Modulation power density of noise at the output of the modulation filter tuned to 25 Hz. Parameter is the bandwidth of the noise.

(ac-coupled) modulation power of the noise within the transfer function of the specific modulation filter, the curves from this figure must be integrated across modulation frequency. This integrated magnitude, the (ac-coupled) modulation power $P = P(f_{mod})$, is proportional to the area under each curve of the figure and is quantitatively given by:

$$P_{\Delta f, \rho, C_{F_{mod}}} = \int_0^{\infty} N_{\Delta f, \rho}(f_{mod}) \cdot W_{C_{F_{mod}}}(f_{mod}) df_{mod} \quad (3.3)$$

Figure 3.10 shows this modulation power as a function of the noise bandwidth for three modulation filters tuned to 5 Hz (\bullet), 25 Hz (\blacklozenge) and 100 Hz (\blacktriangle). In addition, for comparison, the modulation power resulting from integration within the transfer range of a modulation lowpass filter as proposed by Viemeister (1979) is plotted in the figure (+). The corresponding curve has a “lowpass” shape

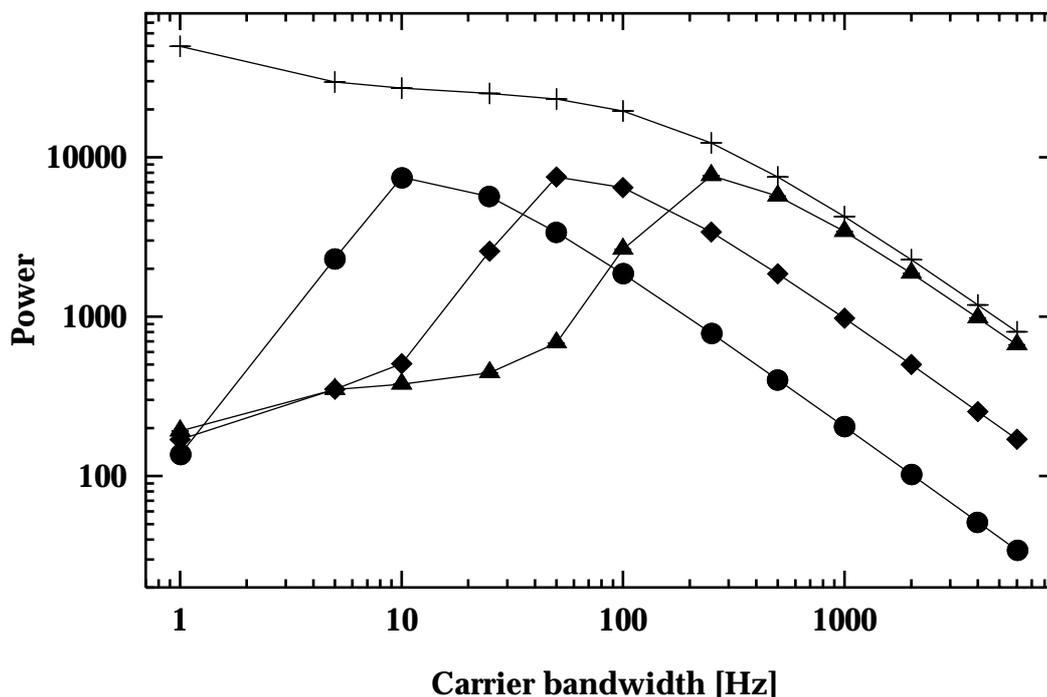


Figure 3.10: Integrated modulation power of noise at the output of the modulation filter tuned to 5, 25, and 100 Hz as a function of noise bandwidth. Parameter is the imposed modulation frequency: $f_{mod} = 5$ Hz (\bullet), $f_{mod} = 25$ Hz (\blacklozenge), $f_{mod} = 100$ Hz (\blacktriangle). In addition, the integrated modulation power at the output of a modulation low-pass filter (Viemeister, 1979) is plotted (+).

whereas the three modulation power functions at the output of the modulation filters have a “bandpass” shape.

In order to compare the shapes of the modulation power curves for the different modulation frequencies, it is necessary to take into account how the corresponding filters attenuate the imposed test modulation. For the filterbank, the attenuation is the same for all three modulation frequencies, because the modulation filter tuned to the test modulation is always used. The modulation lowpass filter, on the other hand, attenuates modulation frequencies according to its transfer characteristic. Viemeister suggested a cut-off frequency of 64 Hz. For such a characteristic, the curves for $f_{mod} = 5$ Hz and $f_{mod} = 25$ Hz are exactly the same whereas the curve for $f_{mod} = 100$ Hz is shifted towards higher values (not shown). The form of the pattern, however, remains the same.

The calculated pattern of the modulation power at the output of the specific modulation filters has a similar form to the threshold pattern from Fig. 3.7. In particular, the peak frequencies in the calculated pattern correspond to those in the measured data. In contrast, the pattern from the calculations based on a modulation lowpass filter, according to Viemeister (1979), differs strongly from

the measured data.

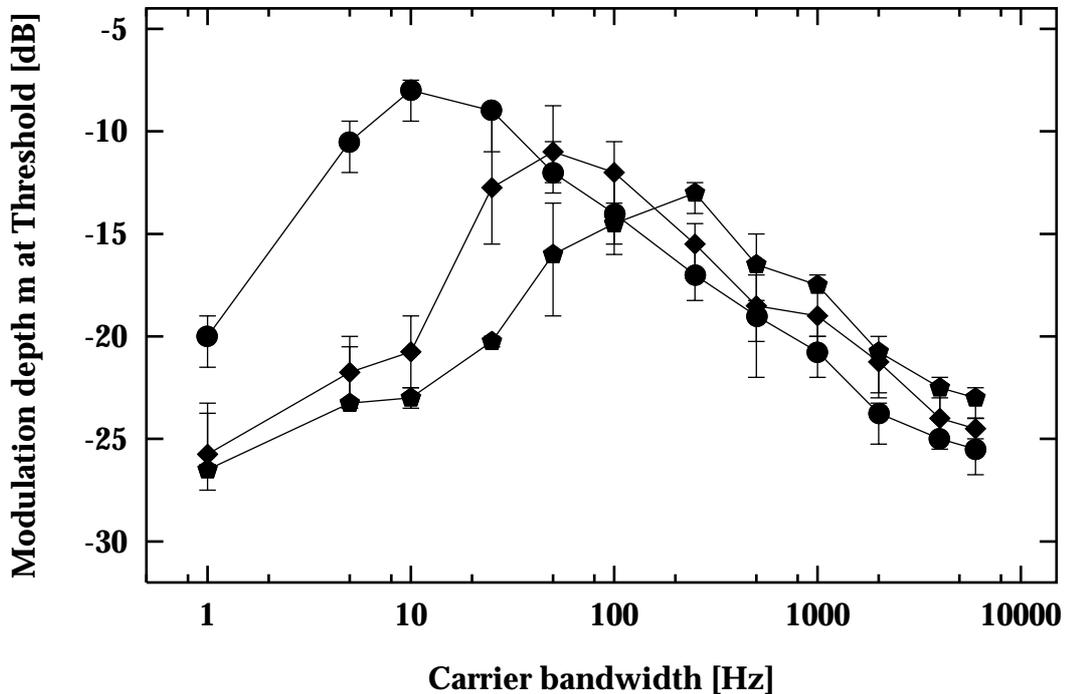


Figure 3.11: Simulated modulation detection thresholds on the basis of the modulation filterbank model, plotted as a function of the carrier bandwidth. Parameter is the signal modulation frequency: $f_{mod} = 5$ Hz (\bullet), $f_{mod} = 25$ Hz (\blacklozenge), $f_{mod} = 100$ Hz (\blacktriangle).

It is questionable if modulation detection thresholds should be directly compared with overall modulation power. In the present and other models of modulation detection for noise carriers, the limiting quantity is the variation in the external stimulus. Thus, it is necessary to estimate not only the amount of modulation power, but also the variation of this value. For a single spectral component in the modulation spectrum of a bandpass Gaussian noise, the power is exponentially distributed (Green et al., 1992). This distribution has the specific property that the mean and standard deviation are equal. If one assumes that the standard deviation in the external noise modulation is the limiting factor in modulation detection, one can expect that the pattern of modulation thresholds runs parallel with the calculated modulation power in Fig. 3.10. This argument holds, however, only as long as one has only *one* modulation spectrum component within the passband of a modulation filter. If the filter passes several spectral components, the distribution of the total power has to be calculated by adding exponential distributions. Without going into mathematical details, it can be stated that the ratio between the standard deviation and mean of the total power will decrease with the number of statistically independent spectral lines. The reason for the

decrease is that means and *variances* add linearly. This statistical effect will lead to somewhat lower thresholds for higher modulation filters. Applied to the three curves in Fig.3.10, which have comparable peak power values, this reasoning predicts modulation thresholds for the peaks which are lower the higher the modulation frequencies is, in accordance with the experimental data in Fig. 3.7.

The point of the analytic considerations was to figure out what essential mechanisms are involved in the modulation detection tasks studied here. To establish realistic detection thresholds it is necessary to include some auditory signal processing and detection device as realized, for example, in the modulation filterbank model. Fig. 3.11 shows corresponding simulated thresholds obtained with the modulation filterbank model. In agreement with the data, the threshold functions have a “bandpass” shape and the peaks of the three curves are lower the higher the modulation frequency is. However, especially for $f_{mod} = 25$ and 100 Hz, the model overestimates human performance in conditions with $\Delta f \geq 1000$ Hz by about 5 dB.

The following two conclusions can be derived from this analysis. Firstly, peripheral filtering seems to play no major role in modulation detection. The whole bandwidth is used by the system. Secondly, it is the interaction between the statistical properties of the applied stimuli with the modulation filters that accounts for the data.

3.4.4 Predictions for modulation masking using broadband noise carriers

Houtgast (1989) adopted a classical masking paradigm for investigating frequency selectivity in the modulation frequency domain: the detectability of a test modulation in the presence of a masker modulation, as a function of the spectral difference between test and masker modulation.

The carrier in all his experiments was a pink noise with a spectrum level in the 1-kHz region of about 25 dB SPL. After applying the modulation, the carrier was bandpass filtered between 1 and 4 kHz and added to unmodulated pink noise with a complementary (bandstop) spectrum. The masking patterns (of the first experiment) were obtained for each of three 1/2-oct-wide bands of masker-modulation noise. The carrier noise was multiplied both with the masker and the target modulator waveform. Center frequencies were 4, 8, or 16 Hz. The left panel of Fig. 3.12 shows the resulting masking patterns from that study. The three curves show a peaked characteristic: the amount of masking decreases for increasing frequency difference between test modulation and masker modulation. The lower curve in the figure shows the unmasked modulation detection threshold level as a function of modulation frequency. For details about the experimental setup, stimuli and procedure, see Houtgast (1989).

The right panel of Fig. 3.12 shows results from simulations obtained with the

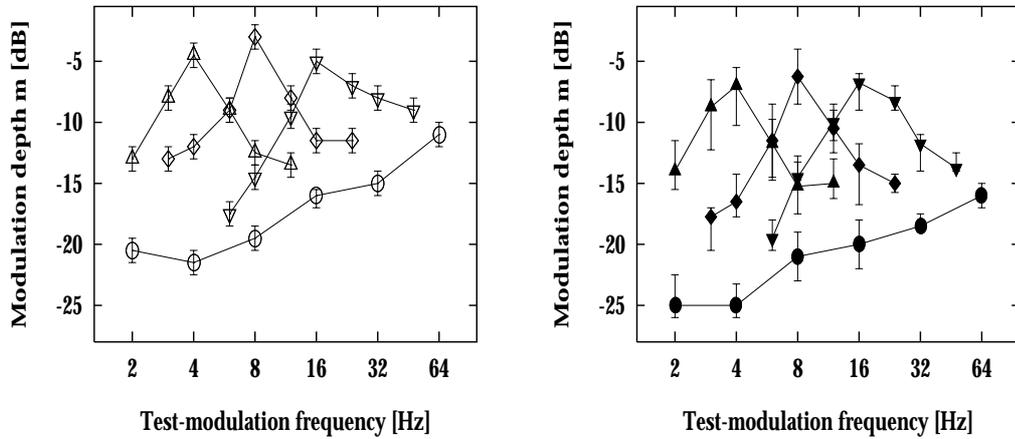


Figure 3.12: The lower curve in both panels shows the unmasked modulation detection thresholds as a function of modulation frequency. Each of the three peaked curves show the masked modulation-threshold pattern for one of the 1/2-oct-wide masker-modulation bands centered at 4 Hz (\triangle), 8 Hz (\diamond), or 16 Hz (∇). The carrier in all measurements was a pink noise ranging from 1 to 4 kHz. The left panel shows data from Houtgast (1989). The right panel shows predictions with the multi-channel model.

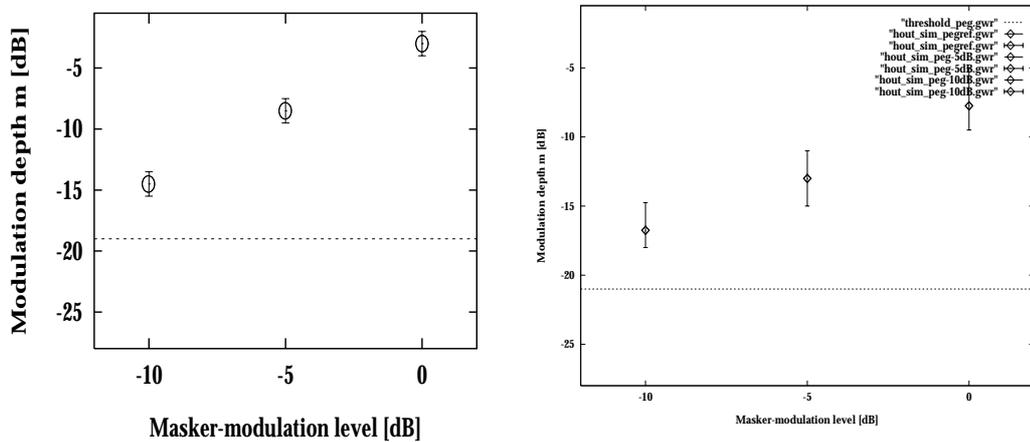


Figure 3.13: Modulation-detection thresholds for an 8-Hz modulation as a function of the relative masker-modulation level for the masker-modulation band centered at 8 Hz. Left panel: Data from Houtgast (1989). Right panel: Corresponding predictions of the modulation filterbank model.

multi-channel version of the modulation filterbank model. The simulated thresholds for the unmasked condition are systematically lower than the experimental data from Houtgast. This might be caused by the different presentation of the masker and by differences in the applied threshold procedure. Houtgast used a continuous masker whereas in the simulations the masker was gated. While in the simulations, a 3-IFC procedure with a 2-down 1-up adaptation rule was used, Houtgast measured thresholds in a 2-IFC procedure with a 3-down 1-up adaptation rule. Besides this systematic difference of about 2-5 dB, both the unmasked thresholds and the masked threshold in both panels agree very well with each other. The simulated masked patterns show the same peaked characteristic, with less masking for increasing frequency difference between test and masker modulation.

An additional observation can be made from the unmasked modulation data. The difference in the experimental thresholds at 4 and 64 Hz, respectively, amounts to about 11 dB. This is much more than observed in corresponding measurements with a white-noise carrier (cf., Figs. 3.3 and 3.4 in the Sect. 3.4.2). Interestingly, a similar difference between white and pink noise is also observed in simulations. When the experimental condition of Houtgast is simulated with a white-noise carrier, thresholds at 4 Hz modulation are the same as for pink noise. The increase between 4 and 64 Hz is, however, smaller for white noise and amounts to about 7 dB. At present there is no simple explanation for this difference.

As in the study of Houtgast (1989), the effect of a reduction of the masker-modulation level was investigated for one of the conditions in Fig. 3.12 (the 8-Hz modulation frequency for the noise band centered at 8 Hz). The masker-modulation level used to obtain the middle threshold function in Fig. 3.12 was reduced by 5 and 10 dB, respectively. The effect on the model prediction is shown in Fig. 3.13. As in the data of Houtgast, the reduction of the threshold level is equal to the reduction of the masker-modulation level. This indicates that the detection threshold is associated with a constant signal-to-noise ratio.

To further test the ability of the modulation filterbank model to account for modulation masking data the effect of varying the bandwidth of the modulation masker was investigated. Thresholds were obtained for a test modulation frequency of 8 Hz for various values of the bandwidth of the masker modulation. The center frequency and the spectral density within the passband were kept constant. Figure 3.14 shows results from the study of Houtgast (1989) (left panel) and model predictions obtained with the present model (right panel).

Houtgast (1989) also found that for small bandwidths, the modulation threshold increases by 3 dB for each doubling of the masker-modulation bandwidth, whereas for large bandwidths, the threshold remains constant. He proposed, that the modulation detection threshold is associated with a constant signal-to-noise ratio within a filter centered at the test-modulation frequency. To a first approximation, indicated by the two straight lines in the left panel of Fig. 3.14, he suggested a width of the modulation filter of about 1/2–1 oct.

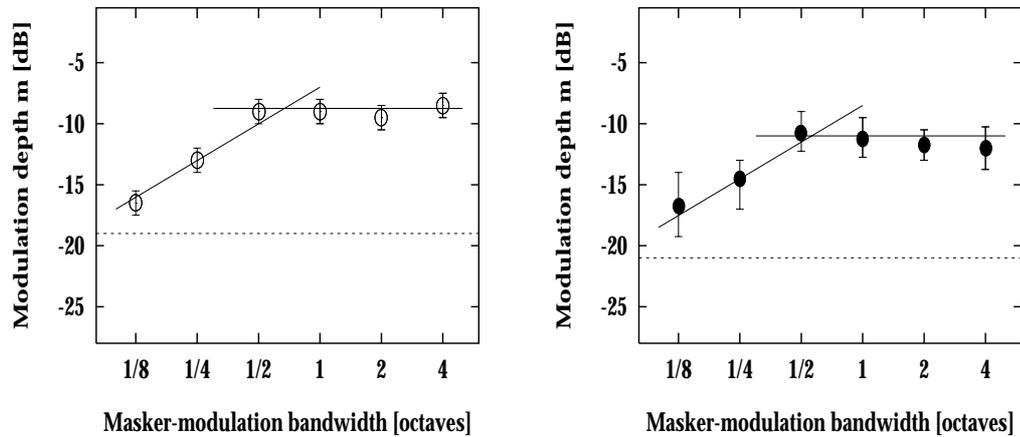


Figure 3.14: Modulation-detection thresholds for an 8-Hz modulation as a function of the bandwidth of the masker-modulation noise centered at 8 Hz. Left panel: Data from Houtgast (1989). Right panel: Corresponding predictions of the modulation filterbank model.

In the present model, the modulation filter centered at 8 Hz has a bandwidth of 5 Hz. This bandwidth lies exactly in the range of 1/2 to 1 octave that was suggested by Houtgast. The model predicts an increase in threshold of about 6 dB between 1/8 and 1/2 octave of modulation-masker bandwidth. For modulation-masker bandwidths larger than a 1/2-oct, it predicts an almost constant threshold. This is in agreement with the data of Houtgast. Thus, modulation masking is only effective within a “critical” band around the test-modulation frequency.

The results from the simulations of the section further support the hypothesis that the envelope fluctuations of the stimuli are processed by modulation-frequency-selective channels.

3.4.5 Temporal integration in modulation detection

Thresholds for detecting sinusoidal amplitude modulation of a wideband noise carrier (lowpass filtered at 6 kHz) were measured and simulated as a function of the duration of the modulating signal. The experimental design was chosen according to the study of Sheft and Yost (1990). The carrier was gated with a duration that exceeded the duration of modulation by the combined stimulus rise and fall times. Stimuli were shaped by a 25.6-ms rise-fall time. The combinations of modulation frequency and numbers of modulation cycles for each condition are listed in Table 3.1. Because modulation was restricted to the constant-amplitude portion of the carrier (thus excluding the ramps), stimulus duration was always 51.2 ms longer than the modulation duration listed in the Table 1 (see Sheft and Yost, 1990).

	Modulation frequency [Hz]								
cycles	2.5	5	10	20	40	80	160	320	640
1	400	200	100	50	25	12.5	6.25		
2		400	200	100	50	25	12.5		
4			400	200	100	50	25	12.5	
8				400	200	100	50	25	
16					400	200	100	50	
32						400	200	100	50
64							400	200	100
128								400	200
256									400

Table 3.1: *Parameter values listing the modulation duration [in ms] for the combination of modulation frequency and number of modulation cycles used in each condition.*

Figure 3.15 show the experimental data of two subjects. The modulation depth at threshold is plotted as a function of the modulation frequency. The parameter is the modulation duration. The data agree very well with those from Sheft and Yost (1990). The curves represent the “classical” broadband TMTF often described in other studies (e.g., Viemeister, 1979; Formby and Muir, 1988). The data show somewhat increased thresholds for the two lowest modulation frequencies, $f_{mod} = 2.5$ and 5 Hz. This is caused by the “gating” of the carrier in this experiment in contrast to experiments where the carrier was presented continuously, as discussed in previous studies (e.g., Viemeister, 1979; Sheft and Yost, 1990). It is assumed that adaptation effects are responsible for this effect, reflecting the temporal relationship between the carrier onset and the onset of

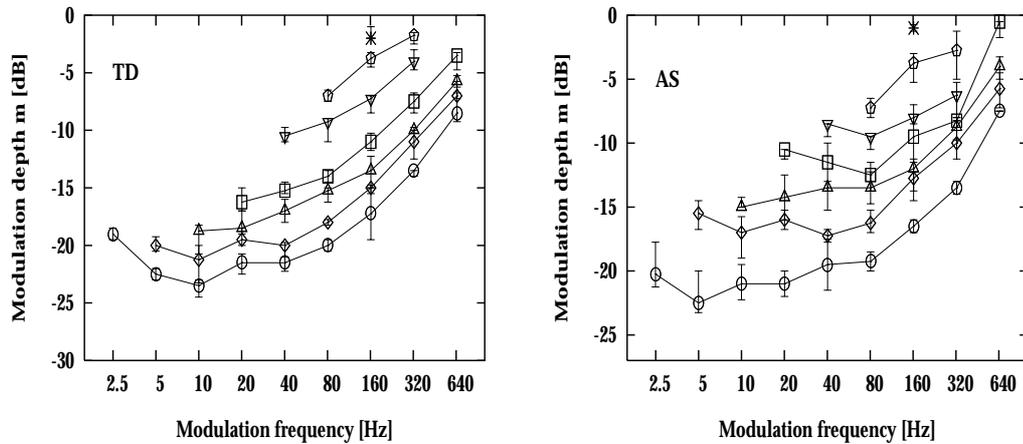


Figure 3.15: Measured TMTF's of two subjects using broadband noise as a carrier. Parameter is the duration of the imposed test modulation. Carrier: 0-6 kHz; Level: 65 dB SPL; Modulation duration: \circ : 400 ms, \diamond : 200 ms, \triangle : 100 ms, \square : 50 ms, ∇ : 25 ms, \circ : 12.5 ms, \times : 6.25 ms; Subjects: TD (left panel), AS (right panel).

modulation. The carrier onset produces a change in the level of the internal representation that is large relative to the changes produced by the sinusoidal modulation and might therefore interfere with the detection of the modulation, especially at low modulation frequencies (Viemeister, 1979). For modulation frequencies larger than 2.5 Hz, the threshold is roughly flat up to a modulation frequency of 40 Hz. It increases slightly at 80 Hz and further increases at a rate of approximately 3 dB/octave for higher modulation frequencies.

Figure 3.16 shows the corresponding simulated thresholds obtained with the modulation filterbank model. The ordinate indicates modulation depth at threshold and the abscissa represents modulation frequency. The simulated threshold pattern is in very good agreement with the pattern found in the measured data. It shows increased thresholds for the two lowest modulation frequencies, 2.5 and 5 Hz, leading to a bandpass characteristic of the threshold function for the two greatest durations of 200 and 400 ms. This is caused by the dynamic properties of the adaption model (Püschel, 1988; Dau et al., 1995a). The feedback loops of the adaptation model produce a considerable overshoot at the carrier onset that decreases the detectability of the signal modulation especially at very low modulation rates (see discussion in Viemeister, 1979; Sheft and Yost, 1990).⁴

⁴Sheft and Yost (1990) found that even when presenting the carrier with a fixed duration that included a 500 ms carrier fringe *preceding* the onset of modulation, threshold patterns exhibit a bandpass characteristic. The adaptation model of Püschel (1988) would not account for such a long-term adaptation effect.

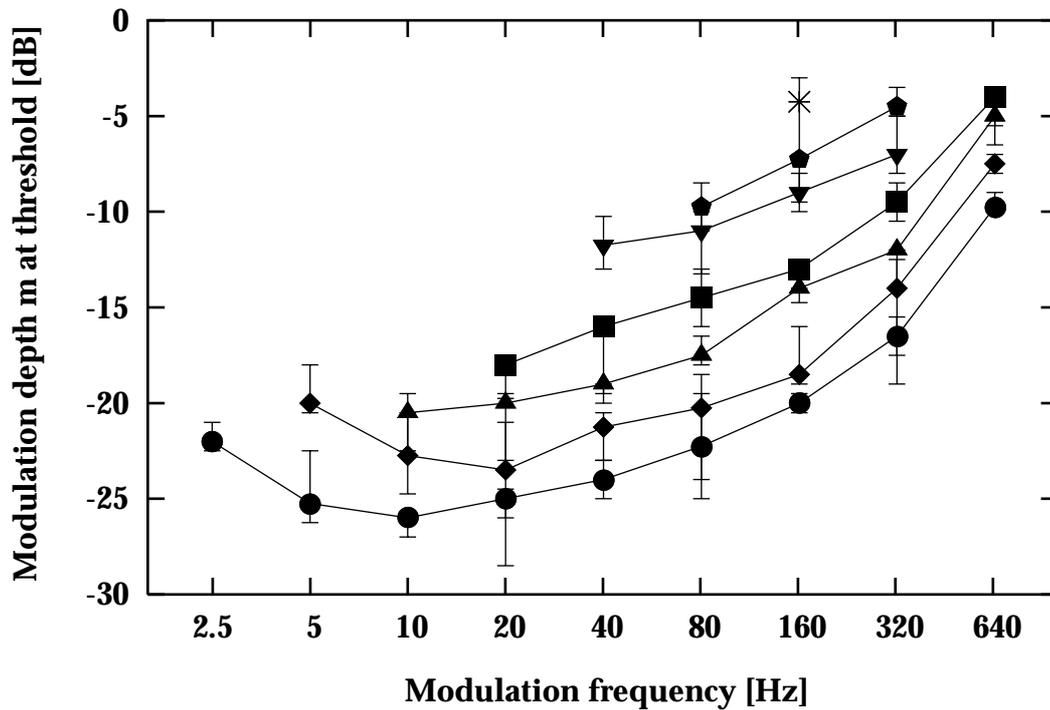


Figure 3.16: *Simulated TMTFs obtained with the current model for the same conditions as in Fig. 3.15.*

The time constants derived from temporal integration per unit duration are in very good agreement with those found in the measured data. The model therefore accounts for the long “effective” integration time constants whose values are much larger than the time constants indicated by the “resolution data” (Sheft and Yost, 1990; Viemeister, 1979).

3.5 Discussion

In this study the performance of the modulation filterbank model with respect to spectral and temporal integration in amplitude modulation detection was described. Several “critical” experiments were performed or taken from the literature and model predictions were compared with experimental data. As an extension to the single-channel model proposed in Dau et al. (1996c), in which the concept of the modulation filterbank was established and the parameters of the modulation filters were fitted to data for narrowband carriers, the modulation analysis was applied to broadband conditions. In this multi-channel version of the model, the temporal envelopes of the stimuli were processed by the same modulation filterbank in parallel at the output of each stimulated peripheral filter. The “observations” at the output of each channel were assumed to be independent of each other. The decision device was realized in such a way that all channels were combined linearly and an optimal weighting of the channels was used.

3.5.1 Spectral integration

In the modulation detection conditions considered in this study using stochastic noise carriers and coherent signal modulation, peripheral filtering generally does not limit the ability to detect modulation. This has been interpreted in some former studies as the observers’ ability to increase their “effective” bandwidth in detecting wideband signals (Green, 1960; Bos and de Boer, 1966; Schacknow and Raab, 1976). As already discussed by Viemeister (1979), the mechanism behind such a combination and the stage of processing where it occurs are still unclear. It has been suggested in the literature (Green and Swets, 1966) that in detection experiments with multicomponent signals, combination probably occurs at a high level: the observer can combine information nearly optimally from widely spaced critical bands. This model approach was adopted within the current study and may be denoted as “auditory filter-based approach”. In contrast, in models for modulation detection conditions, it has been previously assumed that the combination of peripheral filter outputs occurs at a very early stage of processing (Viemeister, 1979). In this context, the model proposed by Viemeister (1979) that includes a wide predetection bandpass filter ($\Delta f = 2000$ Hz) followed by a nonlinearity and a lowpass filter, can account for modulation detection data using (broadband) noise as the carrier. This model approach may be denoted as “predetection-filter approach”. Using broadband carriers, it is not possible to discriminate between such a predetection-filter approach and the auditory filter based approach. Both approaches are mathematically equivalent for broadband carriers, as long as modulation is processed linearly within the model. This is the case for the present model, as described in Dau et al. (1996a,c). Similarly, it is not possible to discriminate whether there is one “large” modulation filterbank behind the combined outputs of *all* peripheral channels (cf., Yost et al., 1989) or,

alternatively, whether there is a modulation filterbank that separately analyzes the output of each peripheral channel before the information is combined. However, the bandwidth of the predetection filter of the Viemeister model is larger than the critical band estimates for most of the auditory range. This approach therefore fails to describe the data in an appropriate way for conditions in which spectral resolution of the auditory system plays any role. The present model gives a more general description of the processing of modulation in the auditory system and is also applicable to spectral-masking conditions, as was shown in Verhey and Dau (1996). Thus, the auditory-filter-based approach has clear advantages over the predetection-filter approach.

One of the main results of this study is that all the data could be described in terms of masking phenomena in the modulation-frequency domain. Analogous to the results described in Dau et al. (1996c), thresholds are mainly determined by the amount of the *inherent* modulation power of the specific noise carrier that falls into the transfer range of the modulation filter tuned to the actual modulation frequency. The model can therefore account for the experimental findings of Eddins that “time constants” derived from the TMTF do not vary with changing frequency region and decrease with increasing stimulus bandwidth. Note, however, that this model does not support the notion of one resolution “time constant” derivable from the data because the lowpass characteristic in the modulation data for broadband noise is not caused by a corresponding “lowpass-weighting” of fast envelope fluctuations by the auditory system. Instead, the data is explained in terms of the “interaction” of statistical properties of the stimuli in the modulation frequency region and the scaling of the modulation filters.

3.5.2 Temporal integration

The phenomenon of temporal integration refers to the fact that, over a range of durations, there is a decrease in threshold with an increase in the duration of the signal. Several models have been suggested in the literature that describe the phenomenon of temporal integration. For example, to describe the threshold function observed in typical test-tone integration data, an integration process is assumed that occurs over relatively long time periods, i.e., of the order of several hundreds of milliseconds. The most prominent approach is the energy detection model (Green, 1960; Green and Swets, 1966) in which decisions are based on the power of the input integrated over a fixed time period. Another class of models assumes a shorter integration time to account for temporal resolution such as modulation detection (Viemeister, 1979), gap detection (Forrest and Green, 1987) and temporal aspects of nonsimultaneous masking (Moore et al., 1988; Oxenham and Moore, 1994). The discrepancy between these two modeling strategies is often described as illustrating the “resolution-integration” paradox (de Boer, 1985).

In recent studies, however, it has been argued that the disparity between the

integration and the resolution time constants is not a real problem (Viemeister and Wakefield, 1991). They pointed out that the observation of a 3-dB decrease in threshold for each doubling of duration - as seen in typical test-tone integration data - means that the auditory system behaves *as if* perfect power integration occurs but that the system is not necessarily performing the operation of mathematical integration. Therefore it might be important to distinguish between the phenomenon of temporal integration and the process that accounts for the phenomenon. Viemeister and Wakefield (1991) provided evidence that integration with a long time constant, such as proposed by the classical models, does not occur in auditory processing. They showed that the threshold for a pair of short pulses yields classic power integration only for pulse separations of less than 5-10 ms. For longer separations, the thresholds did not change with separation and the pulses appeared to be processed independently (cf. Zwislocki et al., 1962). In a second experiment, Viemeister and Wakefield (1991) showed that the threshold for a pair of tone pulses was lower than for a single pulse, indicating some type of integration, but was not affected by level changes of the noise which was presented between the two pulses. The experimental results from that study are plotted in the left panel of Fig. 3.17. It shows the average thresholds for the first pulse alone (squares), the second pulse alone (circles), and for the pulse pair (triangles) as a function of the relative level of the intervening noise. The thresholds for the first pulse alone do not depend on the noise level. There is a slight increase in threshold for the second pulse reflecting forward masking. The thresholds for the pulse pair are about 2.5 dB lower than those for either pulse alone and do not depend on the level of the intervening noise (for details see Viemeister and Wakefield, 1991). These data cannot be explained by long-term integration.

Furthermore, as discussed by Viemeister and Wakefield (1991), the results of this second experiment are also inconsistent with the model proposed by Penner (1978). Penner showed that a compressive nonlinearity followed by a short-time constant integration can result in long *effective* integration. However, such a model - as an example of the class of "single-look" integration models - would predict a certain change in threshold depending on the energy of the noise between the two pulses, since the lower threshold for a pulse pair compared to that for the single pulse requires integration at least over the time of separation of the two pulses.

To account for the data, Viemeister and Wakefield (1991) proposed a "multiple-look" model. With such a model, "looks" or samples from a short-time-constant process are stored in memory and can be processed "selectively", depending on the task. The short time constant allows the model, in principle, to account for temporal resolution data. The combination of a short time constant and selective processing allows the model to also account for the data from the pulse pair experiment described above. The authors suggested that in temporal integration tasks, the long effective time constants may result from the combination of information from different looks, not from true long-term integration. However,

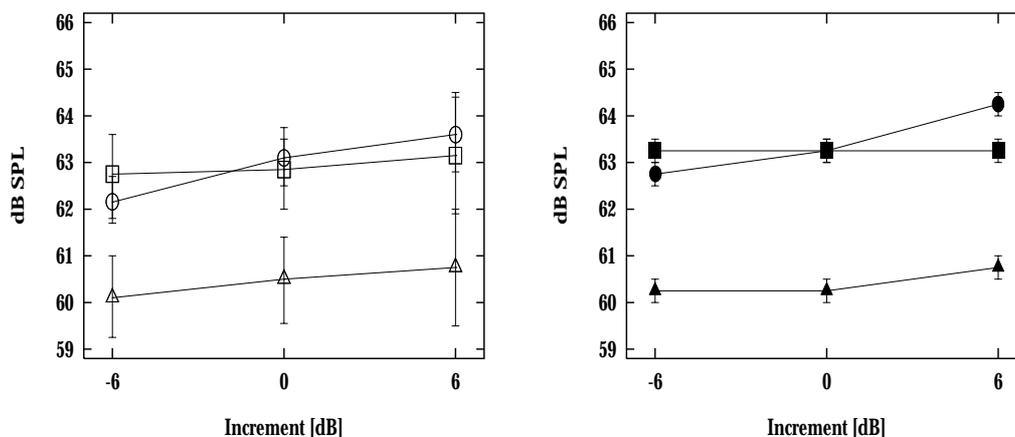


Figure 3.17: *Left panel: Data from Viemeister and Wakefield (1991): Average thresholds as a function of the relative level of the intervening noise. Squares: first pulse only; circles: second pulse only; triangles: pulse pair. Right panel: Corresponding simulated thresholds on the basis of the present model.*

there are some open questions with regard to the multiple-look model proposed by Viemeister and Wakefield (1991). A very basic question is concerned with what is meant by a “look”. The authors discuss the question whether a look may be considered as a sample from the envelope-like waveform from the short temporal window that defines a look. But what is the shape of such a window? What about the correlation between successive looks or samples and how are these looks or samples combined to arrive at a decision statistics? The predicted temporal integration function described in the study of Viemeister and Wakefield (1991) depends strongly on how the looks are weighted and combined.

Nevertheless, the basic concept of the multiple-looks approach is to take into account that the observer attempts to use *all* the samples from the observation interval. For the detection of a tone, for example, an increase in the duration of the tone increases the number of samples and results in an improvement in performance. The model proposed in the present study contains an optimal detector as a decision device. The detection process can be considered as a “matched filtering” process as already described in Dau et al. (1996a). This implies that a *variable* time constant is available that is matched to the signal duration, dependent on the specific task, i.e., the model has at its disposal a continuum of time constants. Time constants that are related to the “hard-wired” part of signal processing within the model represent a lower limit of temporal acuity. (The term “hard-wired” is used in the sense that this part of signal processing is assumed to be independent of the specific experimental task.) The modulation filterbank represents a set of time constants that are, however, too short to account for

the integration data. Thus, it is the decision device that inherently accounts for the long effective time constants. This agrees well with the considerations by Viemeister and Wakefield (1991) that different strategies are probably being employed, and different capabilities tapped, in resolution and integration tasks. The resolution task seems to use more “peripheral” processes whereas temporal integration may require higher-level processes such as multiple sampling and probability summation. To that extent, assuming that the decision process occurs at a higher level of auditory processing, there is a strong correspondence between the modeling strategy of the present model and that suggested in the “multiple-look” model.

This correspondence is supported by the model predictions for the experiment of Viemeister and Wakefield, which are shown in the right panel of Fig. 3.17. Our model predicts a 3-dB decrease in threshold for the pulse pair compared to the threshold for a single pulse. Such a result is expected if the psychometric function for detection of pure tones in noise and in quiet can be described by: $d' = m(\frac{E}{N_0})^k$, where d' is the normalized sensitivity index, E is signal power, N_0 is the spectral noise power density and k and m represent “individual” parameters. For *ideal* observers, the parameter k has the value 0.5 (cf. Green and Swets, 1966), leading to a 3-dB decrease in threshold for the pulse pair compared to that for a single pulse. According to Egan et al. (1969), the psychometric function for normal hearing subjects can be described more accurately with a value $k \approx 1$ instead of 0.5. Such a value would lead to a 1.5-dB instead of 3-dB lower threshold for the pulse pair than for a single pulse, as was discussed by Viemeister and Wakefield (1991).

Long “effective” time constants occur both in typical test-tone integration and in the modulation integration examined in the present study. Whereas the decision device is responsible for the *shift* of the TMTF with changing signal duration, from which an *integration time constant* can be derived, it is the scaling of the modulation filterbank that determines the *form* of the TMTF, from which a *resolution time constant* is commonly derived. As discussed in the section about spectral integration in modulation detection, the threshold is determined mainly by the portion of the modulation power of the broadband noise carrier that is processed by the modulation filters tuned to the signal modulation frequency.

3.5.3 Future extensions of the model

Generally, all predicted thresholds shown in this paper lie between 2 and 5 dB lower than the experimental data. This shift indicates that there must be a certain loss of information in the auditory processing of modulation (independent of modulation rate) that is not at present accounted for in the model. Further modeling efforts are required to understand this discrepancy, that can not be explained by simply increasing the variance of the internal noise, since thresholds are mainly determined by the external statistics of the stimuli.

The present model does not cover conditions which require some processes of *across-channel comparison*. Such across-critical-band processing of temporally modulated complex stimuli might occur in, for instance, comodulation masking release (CMR) (Hall, 1987), comodulation detection difference (CDD) (McFadden, 1987; Cohen and Schubert, 1987), and modulation detection interference (MDI) (Yost and Sheft, 1989). In these cases the auditory system seems to be looking across frequency channels that contain temporally modulated stimuli.

In order to predict performance for these types of experiments (as well as binaural masking experiments), additional stages have to be included that calculate, for instance, the correlation between the envelopes of different frequency regions (or the two ears). Such stages have not been systematically evaluated within the framework of this study (for a first result on binaural masking, see Holube et al., 1995a,b). However, using the present model as a preprocessing circuit, it might be easier to find realistic across-channel processing stages that also allow correct prediction for conditions in which across-channel processing is *not* needed.

3.6 Conclusions

The multi-channel modulation-filterbank model described in this study can predict a wide variety of experimental conditions, including spectral and temporal integration of modulation detection and modulation masking with broadband carriers.

- Spectral integration is accounted for by combining the independent detection cues from all auditory filters with an optimal decision statistic.
- Temporal integration is accounted for by the variable length of the template that forms the basis of the optimal detector incorporated in the model.
- The combination of the modulation filterbank concept and the optimal decision algorithm presents a powerful strategy for describing modulation detection phenomena in narrowband and broadband conditions.

Chapter 4

Amplitude modulation detection with sinusoidal carriers at different frequencies¹

Abstract

For a set of carrier frequencies (2, 3, 4, 5, 9 kHz), the temporal modulation transfer function (TMTF) was measured in the range of modulation frequencies from 10–800 Hz. These conditions include the range of modulation frequencies in which the cue for modulation detection changes from a temporal (at low modulation rates) to a spectral (at higher rates) cue. The data of all subjects show a flat threshold function up to 100 Hz, independent of the carrier frequency. For modulation rates larger than 100 Hz, detection thresholds increase at a rate of about 2-5 dB/oct., depending on center frequency, and finally decrease when the detection cue changes to detection of the spectral sidebands of the modulation. The flat threshold function up to a modulation of about 100 Hz contradicts the hypothesis of a modulation lowpass filter at about 50-60 Hz as often discussed in the literature.

Results from simulations obtained with the modulation filterbank model were compared with the data. The model which has successfully been applied to modulation detection conditions with narrowband and broadband noise carriers, accounts for the flat threshold up to a modulation rate of 100 Hz and also accounts for the frequency-dependent roll-off in the threshold function.

However, the model does not account for the increase of experimental thresholds for modulation rates above 100 Hz (for high center frequencies). A certain loss of information seems to occur in the “central” auditory processing of high

¹Parts of this chapter are presented in a paper in preparation for J. Acoust. Soc. Am. , written by Ralf Fassel, Armin Kohlrausch and me, with the title “The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers.”

modulation rates which is not included in the present model. Further experiments and modeling efforts are required to clarify the processing of fast modulations (> 100 Hz) in the auditory system.

4.1 Introduction

The sensitivity of the human auditory system to periodic amplitude variations has been tested for different carrier types. The most common carrier is a broadband noise (Viemeister, 1979). Using noise as a carrier has the advantage that only temporal cues can be exploited by the subjects, because spectral cues associated with the sidebands produced by the modulation fall within the carrier and are therefore masked. It is assumed that in general short-term spectral cues are not being used by the subject (Viemeister, 1979; Burns and Viemeister, 1981). The disadvantage is, of course, that these experiments cannot reveal the influence of the carrier frequency on modulation sensitivity. In order to investigate this influence, the carrier can be a narrowband noise or a sinusoid. The experiments and model simulations discussed in this chapter deal with the latter paradigm.

Previous experiments with sinusoidal carriers have revealed the following results (Zwicker, 1952; Viemeister, 1979): From very low modulation frequencies, detection thresholds decrease slightly, reaching a minimum around 4 Hz, increase with increasing modulation frequency and finally decrease as soon as the detection cue changes to detection of the spectral sidebands of the modulation (Zwicker and Fastl, 1990). For example, for a 1-kHz carrier at 50 dB SPL, thresholds decrease from 1–4 Hz, followed by a 3 dB/oct. increase up to modulation frequencies of approximately 60 Hz (Zwicker and Fastl, 1990; Viemeister, 1979). The frequency at the maximum of the increasing part in the temporal modulation transfer function, (TMTF) as well as its threshold value, increase with increasing carrier frequency. The roll-off above the maximum is assumed to be caused by the audibility of sidebands, so that one listens for additional tones, whereas for lower modulation frequencies temporal attributes like fluctuations in loudness or roughness are perceived (e.g., Sek and Moore, 1994).

An apparent problem in analyzing modulation detection thresholds for sinusoidal carriers is the possibility of off-frequency listening. It has been argued that, even below the modulation frequency at which the sidebands are resolved, the observer might increase the effective modulation depth by positioning his/her internal filter away from the carrier frequency in such a way that the amplitudes of the carrier and one of the sidebands are less different (e.g., Viemeister, 1979). However, no model calculations have been published so far that allow estimation of the quantitative influence of off-frequency listening on the audibility of high modulation frequencies.

The present study examines amplitude modulation detection for a set of carrier frequencies (at 2, 3, 4, 5, and 9 kHz) for modulation frequencies in the range from 10 to 800 Hz. The study is mainly concerned with the following question: Is the modulation filterbank model, which has so far been successfully tested in conditions with random-noise maskers, also able to account for modulation detection data with sinusoidal carriers, and, apart from the question of the appropriate “effective” signal processing along the auditory pathway, what is the

detection strategy used in modulation detection? The goal is to describe the TMTF's in terms of frequency selectivity in both the audio-frequency domain and the modulation-frequency domain.

4.2 Method

4.2.1 Procedure and Subjects

Amplitude modulation thresholds for sinusoidal carriers were measured and simulated using an adaptive 3-Interval Forced-choice (3IFC) procedure. The unmodulated carrier was presented in two of three consecutive intervals separated by silent intervals of 300 ms. In the third, randomly chosen interval the carrier was sinusoidally modulated. The subject's task was to specify the interval containing the test modulation. During a threshold run, the modulation depth in dB ($20 \log m$), was adjusted by a 2-down 1-up algorithm (Levitt, 1971) which provides an estimate of the modulation depth necessary for 70.7% correct responses. The step size was 4 dB at the start of a run and was divided by 2 after every two reversals of the signal level until the step size reached a minimum of 1 dB, at which time it was fixed. Using this 1-dB step size, ten reversals were obtained and the median value of the signal levels at these 10 reversals was used as the threshold value. The subjects received visual feedback during the measurements. The procedure was repeated four times for each signal configuration and subject. All figures show the median and interquartile ranges based on four single measurements. All five subjects had experience in psychoacoustic measurements and had normal hearing. They were between 23 and 29 years old.

4.2.2 Apparatus and stimuli

All acoustic stimuli were digitally generated at a sampling frequency of 30 kHz. The stimuli were transformed to analog signals with a two-channel 16-bit D/A converter, attenuated, lowpass filtered at 10 kHz and diotically presented via headphones (HDA 200) in a soundproof booth. Signal generation and presentation were controlled by a SUN-Workstation using a signal-processing software package developed at the psychoacoustic laboratory in Göttingen.

Several experiments on modulation detection were performed. Amplitude modulation thresholds for sinusoidal carriers were measured for modulation rates in the range from 10-800 Hz. The duration of the carrier was 800 ms with 150-ms Hanning ramps. The amplitude modulation occurred in the 500 ms constant part of the carrier and was windowed with a Hanning window over its total duration. The level of the stimuli was 65 dB SPL.

4.3 Experimental results and model predictions

4.3.1 Amplitude modulation detection thresholds for a carrier frequency of 5 kHz

There is an “artificial” upper limit in determining the TMTF, which is imposed by the resolution of the sidebands. The influence of this limitation can be reduced by using high carrier frequencies, because at high frequencies, the auditory-filter bandwidth is large. Therefore, to allow predictions about temporal processing over a relatively wide range of modulation frequencies, the first experiment deals with modulation detection using a carrier with a high frequency (5 kHz). Figure 4.1 shows modulation detection thresholds as a function of modulation frequency.

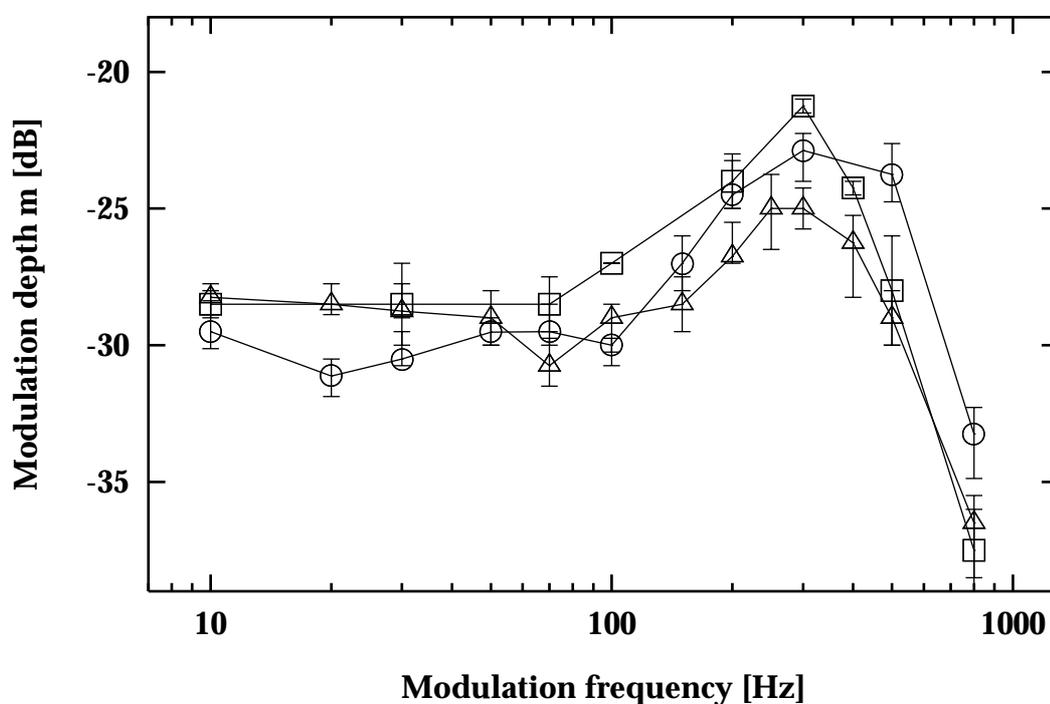


Figure 4.1: Modulation detection thresholds of sinusoidal amplitude modulation as a function of modulation frequency. The carrier was a sinusoid at 5 kHz. Carrier duration: 800 ms; Modulation duration: 500 ms; Level: 65 dB SPL. Subjects: JV (\triangle), PD (\square), TD (\circ).

quency for three subjects. The threshold functions exhibit *three* different characteristic regions. In the range from 10–100 Hz all subjects show a more or less flat threshold function. The threshold values differ from $m = -27$ to -31 dB across subjects. In the range from 100–400 Hz threshold increases at approximately

3-4 dB/oct. Finally, for modulation frequencies larger than 400 Hz a roll-off in the threshold function can be observed; the detection threshold decreases.

The observed threshold function differs from the results of Zwicker (1952). His data show an increase of the threshold values for modulation frequencies above 10 Hz. On the other hand, the results agree quite well with more recent data from Fleischer (1982) and Fassel (1994). Both studies showed modulation thresholds for a 5-kHz sinusoidal carrier to be flat up to modulation frequencies of 100 to 130 Hz. The data of Fassel also show that the same result is obtained at a carrier frequency of 10 kHz. This flat threshold function up to a modulation frequency of 100 Hz contradicts the hypothesis of a modulation lowpass filter in the auditory pathway with a fixed cut-off frequency at about 64 Hz as discussed in the literature (e.g., Viemeister, 1979). If one wishes to keep the hypothesis of such a lowpass filter, its cut-off frequency would have to be at about 150-200 Hz at this carrier frequency.

4.3.2 Comparison of sideband detection and amplitude modulation detection data

A tone-on-tone masking experiment was performed to investigate the extent to which each sideband of the modulated signal contributes to the modulation detection threshold. The (masked) thresholds of the lower and the upper sideband were measured separately in the presence of the carrier as masker (cf. Sek and Moore, 1994). Measurement were performed as a function of the spectral distance Δf between the sideband and the carrier.

Figure 4.2 shows data for three subjects. The two upper panels and the lower left panel of the figure show individual data and the lower right panel shows the averaged data of all subjects. The ordinate indicates signal level at threshold and the abscissa represents Δf . In addition, modulation detection thresholds are included in each panel. Modulation depth at threshold was transformed into signal level per sideband. In the case of the modulation detection condition, the abscissa indicates modulation frequency.

Several interesting effects can be observed in the data. First, for the largest measured spectral distance of 800 Hz, the threshold for the lower sideband for all subjects equals the modulation detection threshold at 800 Hz. This indicates that, at such a high modulation frequency, modulation detection threshold is determined by the masked threshold of the spectrally resolved lower sideband. Second, in the range from 10–100 Hz spectral distance between sideband and carrier, the masked threshold for both the upper and the lower sidebands is almost the same. This result suggests that in the modulation detection condition where both sidebands are presented simultaneously and with equal level, the sidebands are weighted equally by the auditory system. Third, in this range from 10–100 Hz, thresholds for single sideband detection are almost exactly 6-dB higher

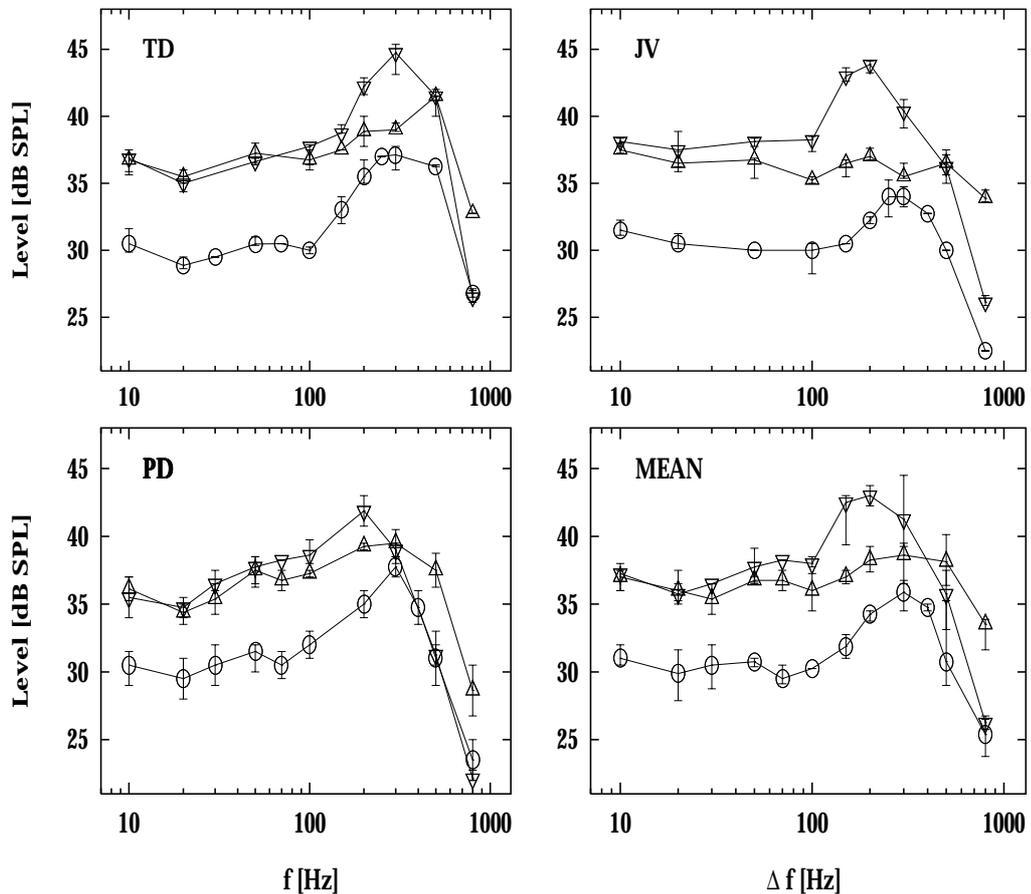


Figure 4.2: Masked thresholds of a sinusoidal test signal as a function of the spectral distance Δf to the 5-kHz masker. Δf was chosen according to the parameter f_{mod} in the modulation detection condition (Fig. 4.1). The figure represents the threshold for the lower sideband (∇) and the upper sideband (\triangle) alone in the presence of the carrier. Masker Level: 65 dB SPL; Masker and test-signal duration: 500 ms. In addition, the modulation detection data from Fig. 4.1 are replotted, where modulation threshold values are transformed into level per sideband (\circ). Subjects: TD (upper left), JV (upper right), PD (lower left), averaged data (lower right).

than the corresponding modulation detection thresholds. If the two sidebands were processed independently in the modulation detection condition (i.e., they were uncorrelated), one would predict a 3-dB difference instead of the observed 6 dB difference (e.g., Green, 1967). The 6-dB difference thus indicates that in the modulation detection task, the upper and lower sideband are *not* processed independently.² For a direct comparison between single-sideband detection and amplitude-modulation detection, one also has to pay attention to differences in envelope *shape* between the two conditions. This difference is very pronounced, if one compares the envelopes of a perfect-beat condition (two sinusoids with equal amplitude) and a fully amplitude modulated (m=1) stimulus. This difference, however, is quite small for modulation depths near threshold and does not play a major role in these conditions. On the basis of these experimental results it can be assumed that for modulation frequencies up to 100 Hz both sidebands are processed within one peripheral channel which is centered at or near the carrier frequency.

An interesting aspect, for which so far no explanation can be given, is the result for the lower sideband between 100 and 300 Hz spectral distance. Phenomenologically, the threshold curve runs parallel to the TMTF curve. This suggests that the detection of the lower sideband is governed by the same cues as in the case of the TMTF. In contrast to the typical asymmetry between lower and upper sideband thresholds at large spectral distances (e.g., 800 Hz), in this region the threshold for the lower sideband is *higher* than that for the upper sideband. One might expect that a combined investigation of lower-sideband audibility and TMTF for sinusoidal carriers will help to better understand the involved mechanism. Such an investigation is, however, beyond the scope of the present study.

4.3.3 Simulations on the basis of the modulation filterbank model

The question remains why modulation detection thresholds increase in the range from 100 to 400 Hz modulation frequency. A possibility is that above 100 Hz the sidebands are increasingly attenuated by the transfer function of the auditory filter (centered at 5 kHz), and that between 100 and 500 Hz modulation frequency the spectral resolution of the auditory system is not good enough to determine the thresholds while modulation depth is already effectively reduced. This question will be addressed in the following two sections.

In the previous chapters a modulation filterbank model was developed, which comprises resonance filters with center frequencies in the range 0–1000 Hz. Within

²Two functions $g(t)$ and $f(t)$ are uncorrelated if the integrated cross product is zero, i.e., $\int_{-\infty}^{+\infty} g(t)f(t)dt = 0$. When this is the case, $\int_{-\infty}^{+\infty} [g(t) + f(t)]^2 dt = \int_{-\infty}^{+\infty} g(t)^2 dt + \int_{-\infty}^{+\infty} f(t)^2 dt$.

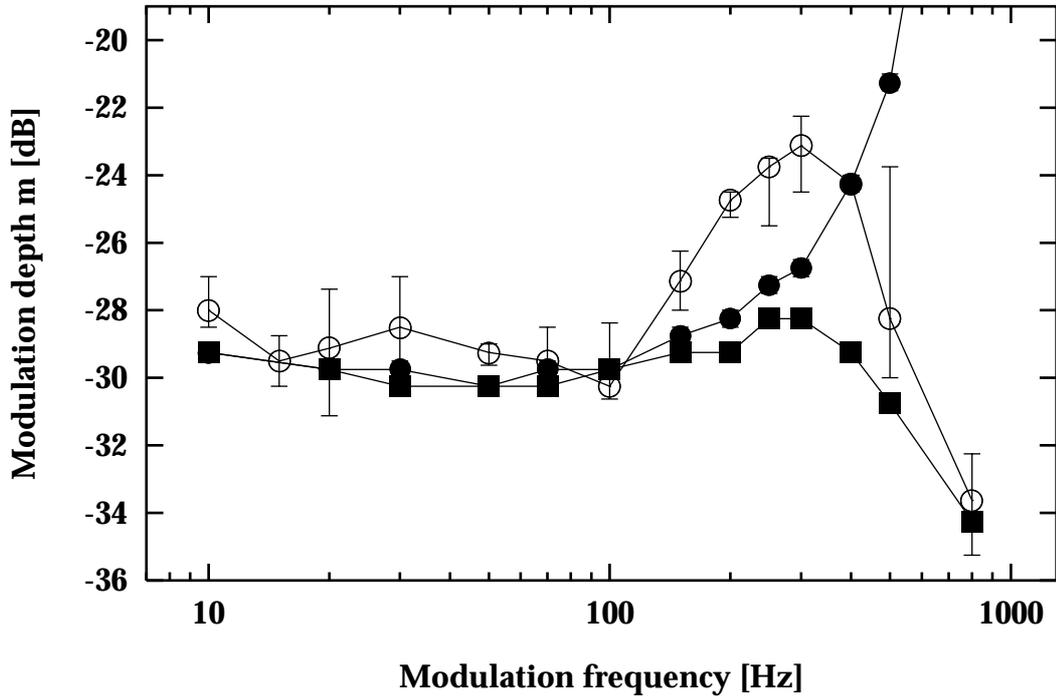


Figure 4.3: Modulation detection thresholds for sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a sinusoid at 5 kHz. The open circles represent averaged data of three subjects (from Fig. 4.1). Simulations are indicated by the filled symbols: The modulation analysis was carried out within the peripheral channel tuned to the carrier frequency (\bullet) and that tuned to the lower sideband (\blacksquare), respectively.

the range 0–100 Hz, the bandwidths of the modulation filters have been determined such that the model accounts for modulation detection data with narrow-band noise carriers. To account for the increase in threshold with a rate of 3 dB per octave of modulation frequency, which is seen in the data with broadband noise carriers, a logarithmic scaling of the modulation filter bandwidth has been chosen: all filters with a center frequency above 10 Hz have a constant Q-value of 2.

The model was developed as a multi-channel model, where modulation filtering was performed across frequency in every excited peripheral channel. In all conditions described in the previous chapters, random noise carriers were used so that it was reasonable to assume *independent* observations across frequency and an increase in information about the signal modulation with increasing number of stimulated auditory channels. The situation is different for a sinusoidal carrier. The limiting factor in this case is the internal noise that, in the present model version, is added as independent noise to the output of all modulation filters in all

peripheral filters. An amplitude-modulated sinusoid generates a coherent modulation pattern in a range of auditory channels. The number of these channels increases with increasing carrier level. Since each peripheral channel contains independent internal noise, one can expect a decrease in modulation threshold with increasing carrier level. In this view, the level of a sinusoidal carrier influences modulation thresholds in a similar way as that for the bandwidth for a noise carrier. As long as the carrier level is fixed, the only difference between a single-channel and a multi-channel analysis is the absolute value of the predicted modulation threshold. It therefore does not seem necessary to apply a broadband analysis to understand the shape of MTFs for sinusoidal carriers. Nevertheless, it is reasonable to make some simple assumptions about the processing of modulation with sinusoidal carriers which can be accounted for by a single-channel model.

Figure 4.3 shows simulations (filled circles) on the basis of the modulation filterbank model where the modulation analysis was first carried out using the peripheral filter tuned to the carrier frequency of 5 kHz. The simulated threshold remains flat in the range from 10–100 Hz, increases slightly in the range from 100–300 Hz and strongly increases for higher modulation frequencies. In the region 10–100 Hz the simulated thresholds agree very well with the averaged data of all subjects which are indicated by the open symbols. In the range from 100–300 Hz, however, the model predicts a weaker increase in the TMTF than observed in the data. With further increase in modulation frequency, the transfer function of the peripheral filter at 5 kHz further attenuates the sidebands of the modulated signal. Therefore, the model predicts an increase in modulation threshold. This is in contrast to the data. As long as the modulation analysis is only performed as a single-channel analysis using a filter centered at the carrier frequency, there is no possibility of accounting for “off-frequency” listening. This may, however, be done by the subjects, at least in conditions in which the sidebands can be spectrally resolved by the system. In such a case, the detection strategy of the subject probably is to monitor one or both auditory filters that are centered near the sidebands of the modulation. For the detection of low modulation frequencies up to 100 Hz, it can be assumed either that no off-frequency occurs or that off-frequency listening is not advantageous for the subject.

A more appropriate decision strategy is required to account for the roll-off in the threshold function for $f_{mod} \geq 400$ Hz. Based on the results from Sect. 4.3.2, it is assumed in the following that the modulation analysis is carried out using a filter centered on the lower sideband. This is motivated by the finding that the modulation threshold for $f_{mod} = 800$ Hz equals the masked threshold of the lower sideband of the signal (see Fig. 4.2).³ The filled squares in Fig. 4.3 show the corresponding simulations. The simulated threshold remains almost flat up

³This does not necessarily mean that the same detection strategy is used by the subjects in these two stimulus conditions. Instead, also in the sideband-detection condition the decision is probably based on off-frequency listening with a maximal signal-to-noise ratio at a lower

to a modulation frequency of 400 Hz, decreases slightly for $f_{mod} = 500$ Hz and predicts the detection threshold for the largest modulation frequency of 800 Hz very well. But, in contrast to the data, there is no increasing part in the threshold function in the range between 100–300 Hz.

4.3.4 TMTFs for different carrier frequencies

To examine the influence of peripheral filtering on modulation detection threshold, additional TMTFs were measured for a set of carrier frequencies. Figure 4.4

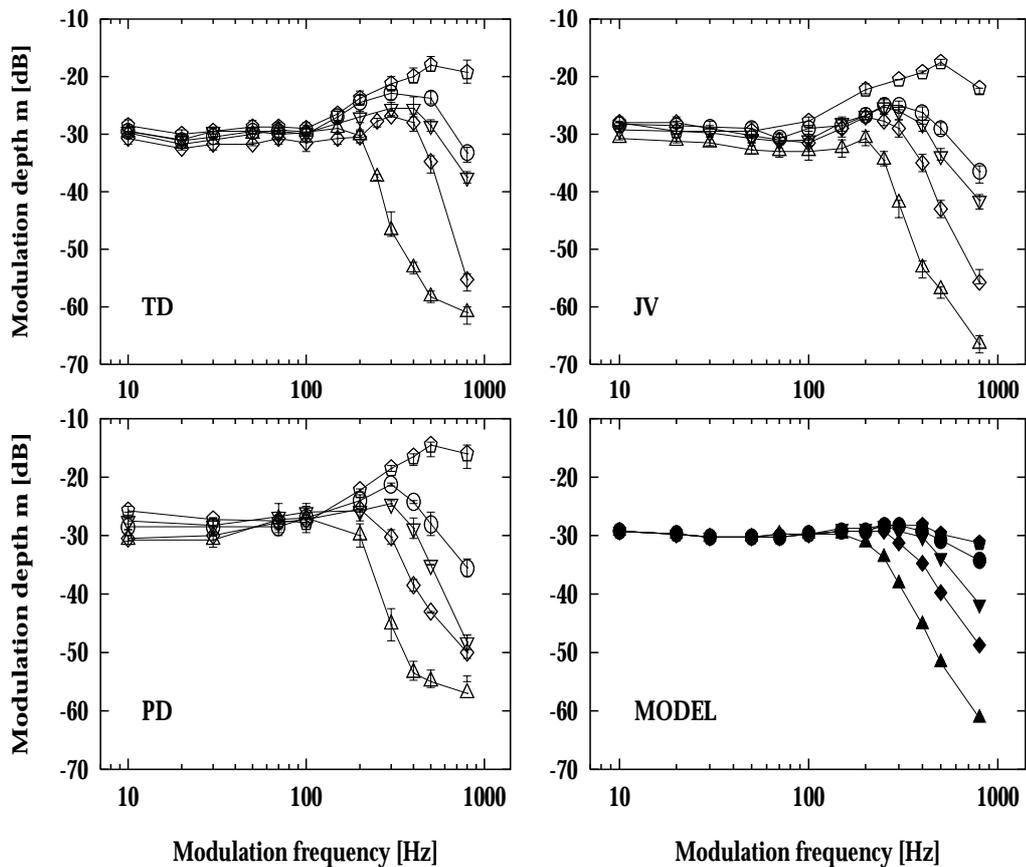


Figure 4.4: TMTFs for carrier frequencies of 2 kHz (\triangle), 3 kHz (\diamond), 4 kHz (∇), 5 kHz (\circ) and 9 kHz (open pentagons). Subjects: TD (upper left), JV (upper right), PD (lower left). Simulated thresholds using the the modulation filterbank model are shown in the lower right panel. The modulation analysis was carried out within the peripheral channel tuned to the lower sideband.

frequency (e.g., Moore, 1986). However, it appears reasonable to assume that in the modulation detection condition thresholds can be estimated by monitoring the filter at the lower sideband.

shows measured and simulated TMTFs for the carrier frequencies of 2, 3, 4, 5 and 9 kHz. The upper two panels and the lower left panel show individual data of three subjects and the lower right panel shows simulations on the basis of the modulation filterbank model where the modulation analysis was performed within the peripheral channel tuned to the lower sideband.

The model accounts for the flat threshold in the range from 10–100 Hz and also accounts for the frequency-dependent roll-off in the TMTF's indicating that, for these regions of modulation frequencies, appropriate detection strategies have been used. However, as already discussed above, the model does not account for the increase in threshold between 100 Hz and the roll-off which can be observed in the experimental data.

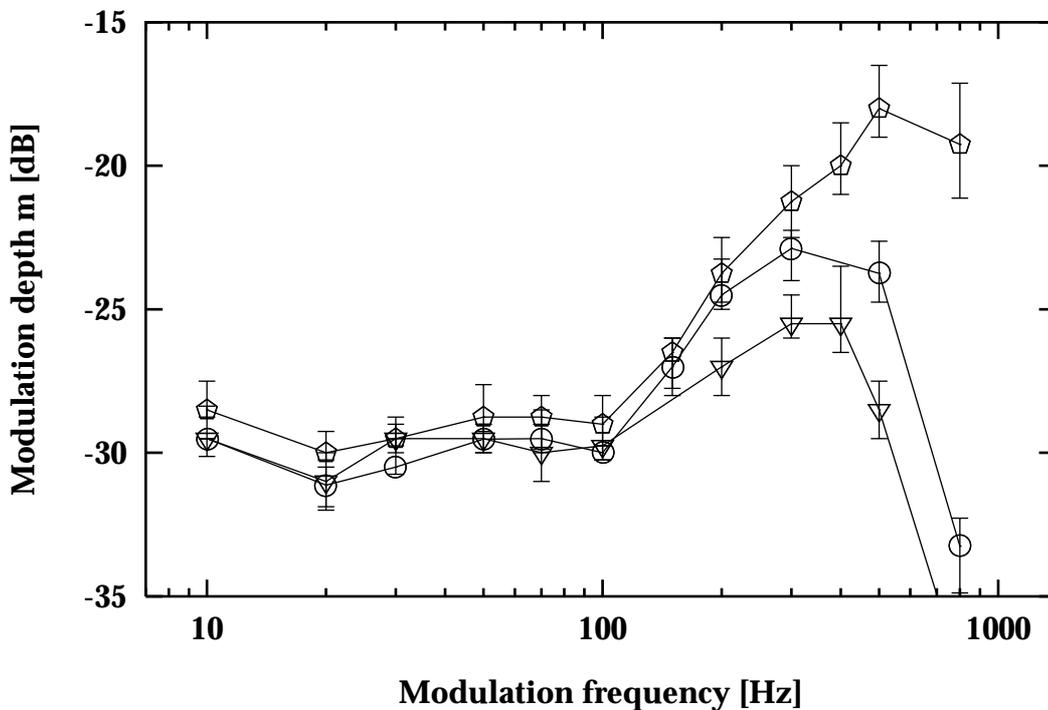


Figure 4.5: *Individual modulation detection thresholds for sinusoidal amplitude modulation as a function of modulation frequency for the carrier frequencies 4 kHz (∇), 5 kHz (\circ) and 9 kHz (open pentagons). Subject: TD.*

To illustrate the increase in threshold in the measured data, Fig. 4.5 shows the individual data of one subject (TD) with a different scale on the ordinate. In this plot, only the results for the three highest carrier frequencies (4, 5, 9 kHz) are shown. At a carrier frequency of 9 kHz, for example, the thresholds increase by about 8–10 dB in the range from 100–500 Hz. Fassel and Püschel (1993) and Fassel (1994) reported very similar results for a carrier frequency of 10 kHz.

If the increasing part of the measured TMTF is determined by peripheral filtering, the frequency at which thresholds *increase* should depend on the carrier frequency because the bandwidth of the peripheral filter increases with increasing center frequency. In particular, thresholds should remain flat up to a higher modulation frequency for the 9-kHz carrier than for 5-kHz carrier. This, however, is not the case. The increasing part of the TMTF's between 100 and 200 Hz is virtually identical for carriers of 5 kHz and 9 kHz. As expected, at a carrier frequency of 9 kHz, the roll-off in the threshold function occurs at a higher modulation frequency compared to that for the 5 kHz carrier.

4.4 Discussion

The present model is able to account well for the roll-off in the data for the highest modulation frequencies. The detection strategy of monitoring the activity in the peripheral filter tuned to the lower sideband seems to be an appropriate approach.

Within the model, there are two effects which determine the slope of the roll-off in the simulated TMTF. The first effect is that, in the filter centered on the lower sideband (holding the level of the sideband constant), the effective modulation depth increases with increasing spectral distance between the lower sideband and the carrier. If we further assume that modulations up to about 1000 Hz are transformed without attenuation within the processing by the modulation filterbank, modulation detection thresholds will decrease for high modulation frequencies. This predicted decrease in threshold is, however, not in line with the common interpretation and observation that the lower sideband can be spectrally resolved at a sufficiently high modulation rate.

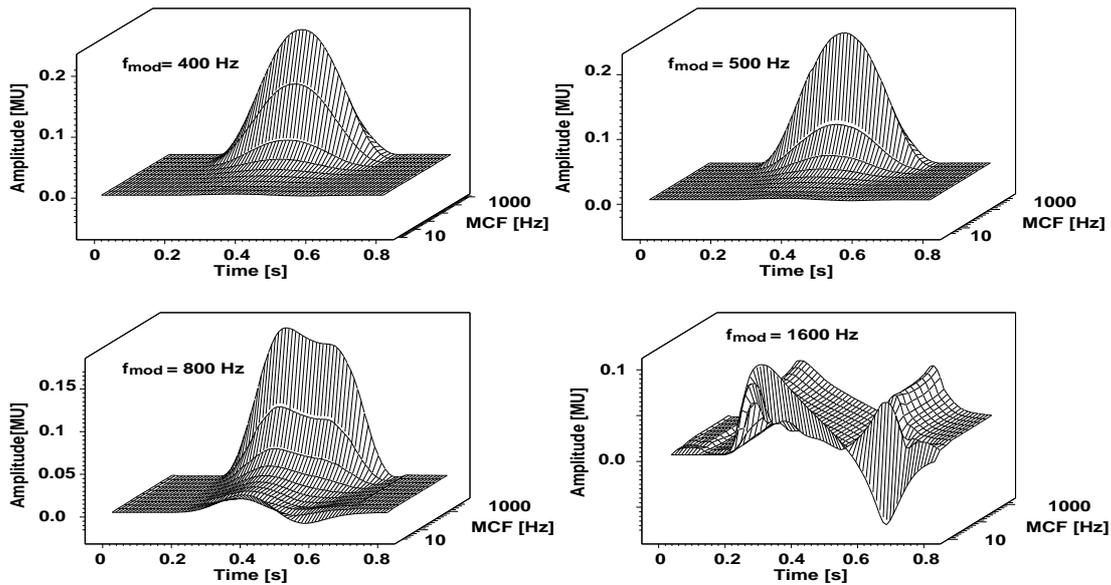


Figure 4.6: *Template-representations for four different stimulus conditions at a carrier frequency of 5 kHz. The z-axis represents modulation center frequency (MCF) of the modulation filters. Upper left: $f_{mod} = 400$ Hz, upper right: $f_{mod} = 500$ Hz, lower left: $f_{mod} = 800$ Hz, lower right: $f_{mod} = 1600$ Hz.*

There is a second effect which determines the threshold in conditions where the sideband is clearly spectrally separated by the system. In such conditions, thresholds are mainly determined by the dc-component of the modulation spectrum, which corresponds to the energy within the auditory filter. This is illus-

trated in Fig. 4.6. It shows derived templates for the modulation rates of 400 Hz (upper left), 500 Hz (upper right), 800 Hz (lower left) and 1600 Hz (lower right) using a carrier frequency of 5 kHz. With increasing f_{mod} there is a systematic shift of the “excitation” from a highly activated modulation filter centered at 400 Hz towards an activated dc-part in case of 1600 Hz. Thus, the simulated threshold depends on the processing of fast modulations up to 800 Hz, whereas at 1600 Hz and higher modulation rates threshold is determined by the excitation of the modulation filter tuned to the lowest modulation frequencies (including the dc-component).

However, the model does not account for all aspects of the data obtained with sinusoidal carriers. For high carrier frequencies, experimental thresholds increase for modulation rates above about 100 Hz, whereas the model predicts nearly constant thresholds. A certain loss of information seems to occur in the auditory processing of high modulation rates which is not included in the present model. Given that the cut-off frequency of around 100 Hz remains constant with center frequency, it can be concluded that the increase in threshold for $f_{mod} > 100$ Hz is not determined by the modulation transfer characteristic of the auditory filter, but that its source has to be at a higher stage of processing.

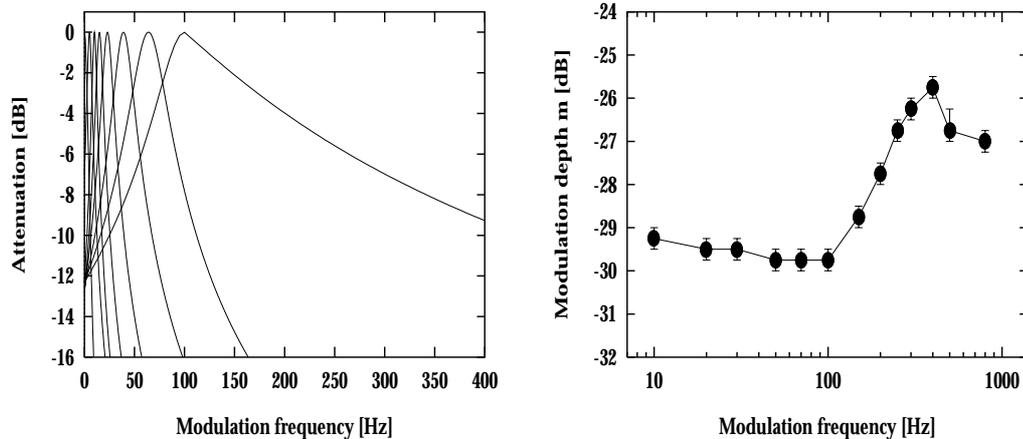


Figure 4.7: *Left panel: Modified modulation filterbank model in which it is assumed that the modulation filter tuned to 100 Hz represents the highest available modulation filter. Right panel: Corresponding simulation of the TMTF for a carrier frequency of 5 kHz. The modulation analysis was performed in the peripheral filter tuned to the lower sideband of the modulated tone.*

If one wants to keep the general structure of the present model, it is necessary to change its processing of modulation rates higher than 100 Hz. In the following section, two possible alternatives and their predictions are described. The first approach is to restrict the range of available modulation filters to center

frequencies up to 100 Hz. In this case, modulation rates above 100 Hz are attenuated by the upper skirt of the 100 Hz modulation filter. The processing of lower modulation rates is assumed to be the same as in the “original” model. Such an approach might be interpreted as an analog in the modulation frequency domain to the “end of cochlea” hypothesis in the periphery (Buus et al., 1986; Zhou, 1995), which states that there is no cochlea filter tuned to the highest audiometric frequencies - the highest available cochlea filter is assumed to be tuned in the range of 14-18 kHz, dependent on the subject. Within the modified modulation filterbank model the transfer characteristic of the highest modulation filter would exhibit an asymmetric shape. The modulation filters of the original model have a slope which is too steep to account for the increase of only about 3-dB/oct. in the measured TMTF. Instead, a shallower slope of the upper skirt (of 6 dB/oct.) of the 100 Hz modulation-filter-transfer function was chosen in order to account for the increase in the measured TMTF. Such a modified modulation filterbank model would also account for the 3 dB increase in threshold for $f_{mod} > 100$ Hz in conditions using broadband noise as the carrier. For a noise carrier, the increase in threshold would then be caused in this alternative model by an *attenuation* of the signal modulation and not any more by the logarithmic scaling of the modulation filter bandwidth. For high modulation frequencies, the modified model version therefore acts as a lowpass filtering stage, with an effective “lowpass cutoff frequency” of about 150 Hz.

The left panel of Fig. 4.7 shows the transfer characteristics of the modified modulation filterbank just described. The right panel of the figure shows, as an example, the simulated TMTF for a sinusoidal carrier of 5 kHz on the basis of this model. The simulated thresholds in the range from 100–400 Hz increase by almost 5 dB which agrees well with the data. The threshold decreases slightly for higher modulation rates. The roll-off in the threshold function is weaker than that observed in the data (see Figs. 4.1 and 4.4). The roll off within the modified model happens as a consequence of the attenuation of fast modulations by the 100-Hz modulation filter. For higher modulation rates, when the energy of the spectrally resolved sideband determines the threshold, the TMTF equals that obtained with the original model (not shown). The discrepancy between experiment and simulation at these high modulation frequencies suggests that the peripheral frequency selectivity is not well modeled by the gamma-tone filter bank. As an example, the gamma-tone filter has a symmetric transfer characteristic on a linear frequency scale. Thus, for very high modulation frequencies, the gammatone filter predicts equal detectability of the lower and the upper side component, in contrast to the results in Fig. 4.2. From model simulations by van der Heijden and Kohlrausch (1994, *Hear. Res.* **80**, 38–52, e.g. Figs. 5 and 7) one can conclude that the gammatone filter considerably underestimates the audibility of the lower side component. Thus, the failure of the present model to correctly predict thresholds at the highest modulation frequencies is most likely caused by the peripheral filter stage.

To preserve the idea of resonance filters for the processing of modulations higher than 100 Hz, a second approach is to introduce a different nonlinearity at the output of these modulation filters. So far, the (Hilbert-) envelopes of the representations at the outputs of the modulations filters were computed. This results in a constant amount of energy of the signal modulation for all modulation frequencies at the corresponding modulation filter output. In order to introduce a

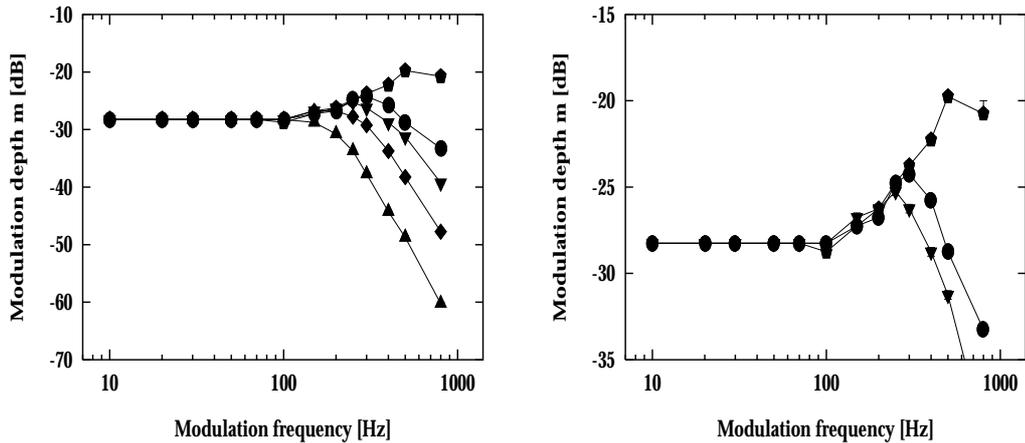


Figure 4.8: *Simulated modulation detection thresholds using a modified modulation filterbank model in which the output of the modulation filters with center frequencies higher than 100 Hz is halfwave rectified and lowpass filtered at 100 Hz (instead of calculating the Hilbert envelope). Parameters as in Fig. 4.4. Left: whole dynamic range of the simulated data; right: simulated thresholds for a reduced dynamic range to better illustrate the increase in threshold for modulation frequencies larger than 100 Hz (only the data for 4, 5, 9 kHz carrier frequency are shown).*

modulation-frequency-dependent information loss, the following modification was tested: instead of calculating the Hilbert envelope, the output of all modulation filters with center frequencies above 100 Hz are half-wave rectified and low-pass filtered at 100 Hz. Assuming the same filter characteristics of the modulation filters as in the original model, such a processing smooths the internal representation at the filter output, and, as a consequence, should lead to an increase in threshold with increasing modulation frequency for $f_{mod} > 100$ Hz.

Such an approach is quite different to a processing by a lowpass filtering (without any prior filtering by a modulation filterbank) where the signal modulation itself would be attenuated according to the transfer function of the lowpass filter. Instead, the *fluctuations* at the output of the modulation filter are smoothed according to the characteristics of the lowpass filter.

Figure 4.8 shows simulated TMTFs for the carrier frequencies 2, 3, 4, 5, and 9 kHz on the basis of the modified modulation filterbank model just described. The thresholds agree very well with the data from Fig. 4.4. The model accounts

for the increasing part of the TMTFs for $f_{mod} > 100$ Hz as well as for the roll-off in the threshold functions.

4.5 Conclusions

In this chapter, the modulation filterbank model, whose parameters had been previously determined in conditions with narrowband noise as the carrier, was tested by measuring modulation detection thresholds for sinusoidal carriers for a range of carrier frequencies.

The model accounts for the flat TMTF up to a modulation frequency of about 100 Hz, independent of the carrier frequency. It also accounts for the threshold of the spectrally resolved sidebands, if the modulation analysis is performed in the peripheral channel tuned to the lower sideband.

The model is not able to describe the increase in threshold observed in the range of modulation frequencies from 100–300 Hz for carrier frequencies larger than 3 kHz, for which the sidebands are not yet spectrally resolved. Therefore, there seems to be a different processing mechanism for modulation frequencies above 100 Hz.

To model the increase in threshold two alternative modeling approaches were discussed. The first approach included a change in the scaling of the modulation filters, and the second approach included a different nonlinearity after modulation filtering. Further experiments are required to understand more completely the processing of high-frequency modulation.

Chapter 5

Summary and conclusion

In this thesis a new modeling approach is developed which is able to predict human performance in a variety of experimental conditions related to modulation detection and modulation masking. The model structure builds on an earlier version of a model of the “effective” signal processing in the auditory system (Dau, 1992; Dau et al., 1995a,b), which allows the prediction of masked thresholds in a variety of simultaneous and non-simultaneous conditions. This “original” version of the model, however, can not deal with conditions in which fast amplitude modulations have to be detected, because the final stage of preprocessing is modeled by a lowpass filter with a cut-off frequency of 8 Hz. In the new version of the model, evaluated here, the lowpass-filter stage is replaced by a modulation filterbank. This modification ensures a certain sensitivity for fast temporal envelope fluctuations and enables the model to describe masking effects in the modulation frequency domain.

The parameters of the modulation filterbank are adjusted to account for modulation detection and modulation masking data with narrowband carriers at a high carrier center frequency. As the bandwidths of these modulated stimuli are smaller than the bandwidth of the stimulated peripheral filter, spectral cues do not play any role and the analysis can be restricted to one auditory filter. In this *single-channel* version of the model, the parameters of the filterbank are divided into two domains with different scaling: in the range 0-10 Hz, the filters have a constant bandwidth of 5 Hz, with the lowest filter being a lowpass filter with a cut-off frequency of 2.5 Hz. Between 10 and 1000 Hz a logarithmic scaling with a constant Q-value of 2 is assumed. This leads to the following predictions: For conditions in which the modulation frequency (f_{mod}) is smaller than half the bandwidth of the carrier (Δf), the model predicts an increase in modulation thresholds with increasing modulation frequency. This prediction agrees with the lowpass characteristic in the threshold functions, which is well known from measurements of the temporal modulation transfer function (TMTF) in the literature. Within the model this lowpass characteristic is caused by the logarithmic scaling of the modulation filter bandwidth. Due to the increasing filter

bandwidth, the power of the intrinsic fluctuations of the noise carrier increases with increasing center frequency of the filters. This increase leads to a threshold increase of approximately 3 dB per doubling of the modulation frequency. In conditions with $f_{mod} > \frac{\Delta f}{2}$, the model can account for the highpass characteristic in the threshold function, reflecting the auditory system's frequency selectivity for modulation. Again, the determining factor is the amount of the inherent modulation power in the carrier falling within the passband of the modulation filter tuned to the test modulation frequency.

In modulation detection conditions with carrier bandwidths larger than a critical band, an extension of the single-channel model is required that allows integration of signal information across frequency. For this purpose, the single-channel model is extended to a *multi-channel* model that is able to describe effects of spectral integration in amplitude modulation detection and masking. The modulation analysis is performed in parallel within each excited peripheral channel, where the parameters of the modulation filterbank are the same for all auditory filters. In the detection stage of the multi-channel model, the outputs of all modulation filters from all excited peripheral channels are combined linearly and with optimal weights.

The model accounts for the findings that, (i), the “time constants” associated with the temporal modulation transfer functions (TMTFs) for bandlimited noise carriers do not vary with carrier center frequency and that, (ii), the time constants associated with the TMTF's decrease monotonically with increasing carrier bandwidth. The model also accounts for data of modulation masking with broadband noise carriers. The predicted masking pattern produced by a narrowband noise along the modulation frequency scale is in very good agreement with results of Houtgast (1989).

With the multi-channel model, an interesting conclusion can be drawn for conditions with stochastic noise carriers and a coherent deterministic amplitude modulation. In such a condition, it is equivalent to detect the modulation of a single noiseband at the output of one wide peripheral filter or to combine the modulation information across an array of bandpass-filtered uncorrelated noises. Thus, for a broadband noise carrier, it is not possible to discriminate between a predetection-filter approach such as proposed by Viemeister (1979) and the approach of linearly combining independent channels as realized in the present model. However, the bandwidth of the predetection filter of the Viemeister model is larger than the critical band estimates for almost all of the auditory range. Such an approach therefore fails to describe masking data in conditions in which spectral resolution of the auditory system plays any role. The present model seems to give a much more general description of the processing of modulation in the auditory system and, at the same time, does not contradict our knowledge of spectral masking.

To integrate information across time, a “multiple-look” strategy is realized within the detection stage of both the single-channel and the multi-channel model.

This strategy allows the model to account for long time constants derived from the data on modulation integration without introducing true long-term integration. Instead, the long “effective” time constants result from the combination of information from different “looks” via multiple sampling and probability summation.

In modulation detection experiments with deterministic carriers (such as sinusoids), the limiting factor for detecting modulation within the model is the *internal* noise that is added as independent noise to the output of all modulation filters in all peripheral filters. In addition, as a further important difference to the conditions with broadband random noise carriers, the shape of the peripheral filters plays a major role in stimulus conditions where the detection is based on the “audibility” of the spectral sidebands of the modulation. The multi-channel model can account for the observed flat modulation detection thresholds up to a modulation rate of about 100 Hz and also for the frequency-dependent roll-off in the threshold function observed in the data for a set of carrier frequencies in the range from 2–9 kHz. However, there are open questions related to the processing of modulations above 100 Hz (for high center frequencies). It appears necessary here to introduce a further loss of information at high modulation rates.

The present model could also be used as a preprocessing circuit for describing experiments which exhibit some kind of across-critical-band interaction of modulation components (e.g., Hall, 1987; Yost and Sheft, 1989). In order to predict performance for these types of experiments, additional stages would have to be included that calculate, for example, the correlation between the envelopes of different frequency regions.

Appendix A

Contributions from Signal detection theory (SDT)

Idea of the “optimal detector”

Signal detection theory (Green and Swets, 1966) is especially concerned with *decision rules* and is closely connected with statistical decision theory. In the following section the fundamental ideas of optimal detection processing will be discussed.

A.1 Formal discussion of the decision problem

Let us first imagine a typical psychoacoustical observation task. In a single observation interval an acoustical stimulus is presented and one of two possible stimulus alternatives is realized. These alternatives are denoted S and N and are also called hypotheses:

S : During the observation interval a signal is added to noise (“signal interval”).

N : During the non-signal interval only noise is presented (“noise interval”).

These different “states of the world” S or N produce specific *events*, i.e., specific representations of the acoustical stimulus in the auditory system. A fixed stimulus does not always cause the same event e ; rather the events are *distributed*. The probability of an event e , given that S has occurred, is denoted by $P(e|S)$ and the probability of the event e , given that N has occurred, is denoted by $P(e|N)$. Because both alternatives may lead to the same event it is generally impossible to get definite information about the “state of the world” from the event.

The observer's task is to make a decision, i.e., S or N , on the basis of the event e . For this decision he needs a *decision criterion*. The value of this criterion depends on the goal of the task.

To maximize the number of correct decisions, as an example, it is reasonable to select that hypothesis which has the larger probability conditional on the given event e . Therefore the decision S will be made if

$$\frac{P(S|e)}{P(N|e)} > 1 \quad . \quad (\text{A.1})$$

Generally these “a-posteriori probabilities” are not given directly. Instead, assumptions about the distributions $P(e|S)$ and $P(e|N)$ are stated.

According to the *Baye's Theorem* the ratio of the a-posteriori probabilities of two hypotheses S and N is given by :

$$\frac{P(S|e)}{P(N|e)} = \frac{P(S) \cdot P(e|S)}{P(N) \cdot P(e|N)} \quad (\text{A.2})$$

The observer selects the hypothesis S if

$$\underbrace{\frac{P(e|S)}{P(e|N)}}_{\text{likelihood ratio}} > \frac{P(N)}{P(S)} =: c \quad (\text{A.3})$$

This *decision rule* contains the quantity $\frac{P(e|S)}{P(e|N)}$. It is called the *likelihood ratio* and is denoted by $l(e)$ (Green and Swets, 1966; Fischer, 1967). The decision criterion is solely determined by the “a-priori” probabilities $P(S)$ and $P(N)$.¹

Of course, the application of such a decision rule does not lead to infallible results. Basing decisions on probabilities implies the possibility of wrong decisions. The decision criterion c divides the likelihood-ratio continuum (also called decision axis) into two ranges dependent on the decision goal and motivation of the observer. The *probability of being correct* concerning the hypothesis S can be computed. It is the probability of the hypothesis S causing an event e that leads to a likelihood ratio $l = l(e)$ larger than the decision criterion c . This is denoted by :

$$P(\text{correct}) = P(l > c|S) \quad (\text{A.4})$$

The decision rule $l(e) > c$ can be substituted by another decision rule that is *monotonic* with l , because only the scale will be transformed but the decisions will remain the same. Then, the decision criterion must be transformed by the same rule.²

¹Other strategies (for example minimizing the number of false decisions) lead to a different decision criterion.

²For a strictly monotonic function f , $l(e) > c$ can be replaced by $f(l(e)) > f(c)$.

In general it can be shown (Green and Swets, 1966) that each decision rule that is based on the likelihood ratio or another quantity which is monotonically related to the likelihood ratio optimizes the decision maker's goal (for example the number of correct decisions). This fact will be useful for later computations of the likelihood ratio when specific assumptions about the underlying distributions are made. To summarize, the *optimal* decision procedure is to compare a quantity that is monotonically related to the likelihood ratio with an appropriate decision criterion.

A.2 The decision problem in an m IFC task

To avoid the problem of determining an *absolute* decision criterion, investigators often use an *m-Interval Forced-Choice* (m IFC) design for psychoacoustical experiments. In forced-choice tasks more than one observation interval is used. In the most general case, with m observation intervals, the signal is presented in one - randomly selected - interval, so there are $m - 1$ intervals with noise alone and a single interval in which the signal is added to the noise. On the assumption of *uncorrelated* intervals, the problem can be reduced to the description of m independent single intervals. So it is assumed that the observer's decision is determined by the likelihood ratios of the events e_i , where the likelihood ratio is computed in the following way:

$$l(e_i) = \frac{P(e_i|S)}{P(e_i|N)} \quad (i = 1, \dots, m) \quad (\text{A.5})$$

$l(e_i)$ has the same definition it had in the analysis of the single-observation procedure (see Eq. (A.3))

The observer's decision rule is to select the interval producing the largest likelihood ratio and to associate it with the signal interval (Green and Swets, 1966). Because this decision rule is based on the *comparison* of likelihoods it is possible to avoid an *absolute* internal criterion used in the single-observation task. The observer will be correct if the likelihood ratio associated with the signal-plus-noise-distribution is actually larger than each likelihood ratio associated with the noise-alone distributions. But of course the noise-alone distribution may produce a larger ratio l , in which case the observer will make a wrong decision.

Consider an 2IFC task and let the sample of likelihood ratio from the signal distribution be called l_s and the sample from the noise distribution be called l_n . If $l_s = k$, then the observer will be correct if $l_n < k$. If the two samples are independent, then for a given k the probability of a correct response in the two-alternative forced-choice task is:

$$P_{2,k}(\text{correct}) = P(l_s = k) \cdot P(l_n < k) \quad (\text{A.6})$$

The total probability of being correct is evaluated by integrating this expression over all possible values of k :

$$\begin{aligned} P_2(\text{correct}) &= \int_{-\infty}^{+\infty} [P(l_s = k) \cdot P(l_n < k)] dk \\ &= \int_{-\infty}^{+\infty} [P(l = k|S) \cdot P(l < k|N)] dk \end{aligned} \quad (\text{A.7})$$

Extension of this argument to m intervals yields the probability of correct decisions in an m IFC task :

$$\begin{aligned} P_m(\text{correct}) &= \int_{-\infty}^{+\infty} [P(l_s = k) \cdot (P(l_n < k))^{m-1}] dk \\ &= \int_{-\infty}^{+\infty} [P(l = k|S) \cdot (P(l < k|N))^{m-1}] dk \end{aligned} \quad (\text{A.8})$$

To calculate the probability of a correct response explicitly, some assumptions about the underlying distributions must be included.

A.3 Justification of the Gaussian assumption and computation of the probability of correct decisions

So far no assumptions have been made about the particular probability distributions of events. There are two reasons for assuming that the underlying distributions of the (noise) events are Gaussian (Green and Swets, 1966). The first - theoretical - aspect is the “central limit theorem”. It states that the sum of n independent, identically distributed random variables approaches a Gaussian distribution for large values of n , irrespective of the particular distribution. On the assumption that the internal noise that limits the resolution arises from the combination of independent sensory events, the central limit theorem may be invoked to justify the assumption of the Gaussian distribution.

The second - practical - reason is that Gaussian distributions are easy to handle in mathematical calculations, e.g. the mean μ and variance σ^2 are additive under linear transformations. Furthermore, the likelihood ratio can easily be computed if the signal is embedded in Gaussian noise.

In the following statistical analysis it is assumed that, due to the internal noise, the representation $e(t)$ of a whole temporal (acoustical) stimulus at the time t_ν is given by the conditional density $P(e(t_\nu)|N) =: P(e_\nu|N)$ if noise alone

is present. It is characterized by $P(e(t_\nu)|S) =: P(e_\nu|S)$ if *signal plus noise* are presented.

With the following simplifying assumptions:

- $P(e_\nu|S)$ has mean s_ν and variance σ^2
- $P(e_\nu|N)$ has mean 0 and also variance σ^2

the probability of a noise-alone-interval evoking a representation e_ν at time t_ν is given by :

$$P(e_\nu|N) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(e_\nu)^2}{2\sigma^2}} \quad (\text{A.9})$$

Given signal plus noise, the density function for the occurrence of the representation e_ν is denoted by :

$$P(e_\nu|S) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(e_\nu - s_\nu)^2}{2\sigma^2}} \quad (\text{A.10})$$

On the assumption that the different samples e_ν are *uncorrelated* with each other, the representation of the whole temporal stimulus is given by the product of the sample densities, for example :

$$P(e|S) = \prod_\nu P(e_\nu|S) \quad (\text{A.11})$$

Now the likelihood ratio can be computed. It maps a whole temporal interval onto a positive real number.

$$l(e(t)) \equiv l(e) = \frac{P(e|S)}{P(e|N)} = \frac{\prod_\nu \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(e_\nu - s_\nu)^2}{2\sigma^2}}}{\prod_\nu \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(e_\nu)^2}{2\sigma^2}}} \quad (\text{A.12})$$

If we take the natural logarithm and simplify this we obtain

$$\Leftrightarrow \ln l(e) = \frac{1}{\sigma^2} \left\{ \underbrace{\sum_\nu e_\nu s_\nu}_{\text{cross correlation}} - \frac{1}{2} \cdot \underbrace{\sum_\nu (s_\nu)^2}_{\text{energy of } s(t)} \right\} \quad (\text{A.13})$$

According to Eq. (A.13), the cross correlation between the *expected* signal $s(t)$ (“template”) and the *received* signal $e(t)$ is a monotonic function of the likelihood ratio. According to the above explanations, the cross correlation is an appropriate quantity for expressing a decision criterion that is optimal in the sense mentioned above.

To be able to compute the probability of correct responses we need to know the distributions of the likelihoods or cross correlations. If we regard the signals

as vectors, the cross correlation is the scalar product of the Gaussian-distributed event $e(t)$ and the deterministic template $s(t)$, so the correlation must be Gaussian, too. If the detector's energy is normalized to 1, the distributions $P(l|S)$ and $P(l|N)$ have the same variance σ^2 as the distributions $P(e|S)$ and $P(e|N)$. The mean of a noise interval is 0, the mean of the signal interval is denoted by d . According to Eq. (A.8) the probability of being correct in an $mIFC$ task is given by :

$$P_m(\text{correct}) = \int_{-\infty}^{+\infty} \phi\left(x - \frac{d}{\sigma}\right) \Phi(x)^{m-1} dx \quad (\text{A.14})$$

with

$$\Phi(x) = \int_{-\infty}^x \phi(x') dx' \quad (\text{A.15})$$

where $\phi(x)$ denotes the density of the standard normal distribution.

Therefore, according to Eq. (A.14), the probability of being correct depends on d , σ^2 and the number m of the observation intervals.

An alternative way of deriving $P_m(\text{correct})$ in Eq. (A.14) is by computing the probability of the cross correlation value associated with the signal being greater than the *maximum* of all $m - 1$ cross correlation values associated with noise alone. Cramer (1946) showed that the maximum of the noise samples has approximately a Gaussian distribution. Since the difference of two Gaussian "random variables" also has a Gaussian distribution, the difference between the sample associated with the signal and the greatest noise sample must also have a Gaussian distribution. Therefore Eq. (A.14) can be approximated by a Gaussian integral with a mean $\mu^*(m)$ and a variance $\sigma^*(m)^2$ (tabulated in Green and Dai, 1991) :

$$P_m(\text{correct}) \approx \int_{-\infty}^{\frac{d - \mu^*(m)}{\sigma^*(m)}} \phi(x) dx = \Phi\left(\frac{d - \mu^*(m)}{\sigma^*(m)}\right) \quad (\text{A.16})$$

In our simulations, this integral must be calculated whenever the value of the actual variable, which is in most cases the signal level, is changed. According to Press et al. (1989) a closely related mathematical integral, the so-called *error-function* $\text{erf}(x)$, is easily computed :

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (\text{A.17})$$

Appendix B

Transformation characteristics of the nonlinear adaptation circuits

For stationary signals an input value I produces a value of $O = \sqrt{I}$ at the output of the first feedback loop (derived from the stationary condition $\frac{I}{O} = O$). For a chain of n loops we obtain an output of $O = \sqrt[n]{I}$. For $n = 5$ this approaches a logarithmic transform. The model allows an input range of 100 dB. The linear amplitude $x_{min} = 10^{-5}$ is introduced as the “absolute” input threshold and is set to 0 dB. The maximum input value is 100 dB, corresponding to a linear amplitude of 1. Thus, in the state of rest with zero input, we get an output value of $\sqrt[5]{10^{-5}} = 0.7$. Consequently the input range of 0 to 100 dB will be transformed to an output range of about 0.3 (= 1 - 0.7). This range of derived output values was linearly mapped to the range of 0 to 100 model units (MU). For a perfect log transformation, a change of 1 MU would correspond to a 1 dB change at the input, irrespective of the input level (cf. Fig. B.1). With five feedback loops we obtain an input-output characteristic, i.e., a transformation of the input signal level (in dB SPL) into model units (MU), as shown in Fig. B.1. There is a small deviation from a straight line which would represent a logarithm transform on this scale. In the case of an input signal at a low level a certain increment of the input level leads to a somewhat smaller increment at the output than in the case of an input signal at a high level. Since we assume a constant sensitivity in terms of output values (MU), this slight deviation from a logarithmic transform predicts a smaller just-noticeable change in level at high input levels compared with low input levels. The actual values for the just-noticeable change in level for input levels of 20, 40, 60, 80 and 100 dB are 1.44, 1.13, 0.95 (compare section I.D.), 0.82 and 0.80 dB, respectively. If the adaptation system is in its state of rest, the first non-zero sample of a signal is not attenuated but will be divided by a very small value f . This value is given by the product of the charging states of

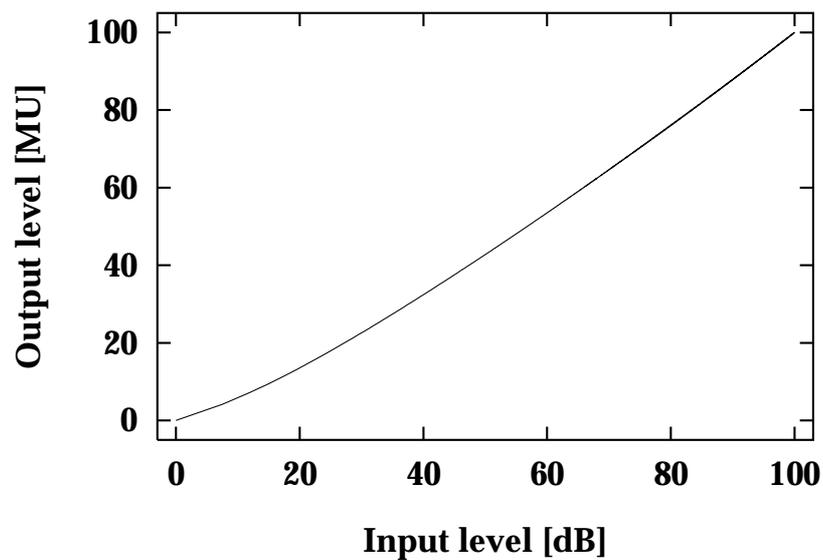


Figure B.1: Transformation of the input level (dB SPL) into the output level (model units MU) for the adaptation model containing five feedback loops.

the capacitors of the different lowpass filters:

$$f = \prod_{n=1}^5 \frac{1}{(10^{-5})^{\frac{1}{2^n}}} = (10^5)^{\frac{31}{32}} \quad (\text{B.1})$$

Thus, the first sample will be multiplied by nearly 10^5 . This leads to an overshoot. After a short time, according to the different time constants, the charging states will increase and therefore the following samples will be attenuated.

References

- Bacon, S.P. and Grantham, D.W. (1989):
“Modulation masking: Effects of modulations frequency, depth and phase,”
J. Acoust. Soc. Am. 85, 2575–2580.
- Bracewell, R.N. (1986):
“The Fourier Transform and Its Applications,”
Second Edition, Revised.
- Bos, C.E. and de Boer, E. (1966):
“Masking and discrimination,”
J. Acoust. Soc. Am. 39, 708–715.
- Burns, E.M. and Viemeister, N. (1981):
“Played again SAM: Further observations on the pitch of amplitude-modulated
noise,”
J. Acoust. Soc. Am. 70, 1655–1660.
- Buus, S., Schorer, E., Florentine, M. and Zwicker, E. (1986):
“Decision rules in detection in simple and complex tones,”
J. Acoust. Soc. Am. 80, 1646–1657.
- Cramer, H. (1946):
Mathematical methods of statistics.
Princeton, NJ: Princeton University Press.
- Creutzfeldt, O.D. Hellweg, F.C. and Schreiner, C.E. (1980):
“Thalamocortical transformation of responses to complex auditory stimuli,”
Exp. Brain Res. 39, 87–104.
- Cohen, M.F. and Schubert, E.T. (1987):
“The effect of cross-spectrum correlation on the detectability of a noise
band,”
J. Acoust. Soc. Am. 81, 721–723.

- Dau, T. and Püschel, D. (**1992**):
 “Computermodell für die Detektion komplexer Schallreize,”
 Fortschritte der Akustik DAGA '92, 877–880.
- Dau, T. and Püschel, D. (**1993**):
 “A quantitative model of the “effective” signal processing in the auditory system,” in *Contributions to psychological acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), 107–120.
- Dau, T., Püschel, D. and Kohlrausch, A. (**1996a**):
 “A quantitative model of the “effective” signal processing in the auditory system: I. Model structure,”
 J. Acoust. Soc. Am. 99, 3615–3622.
- Dau, T., Püschel, D. and Kohlrausch, A. (**1996b**):
 “A quantitative model of the “effective” signal processing in the auditory system: II. Simulations and measurements,”
 J. Acoust. Soc. Am. 99, 3623–3631.
- Dau, T., Kollmeier, B. and Kohlrausch, A. (**1995**):
 “Modeling modulation perception: modulation low-pass filter or modulation filterbank?,”
 J. Acoust. Soc. Am. 97, 3273.
- Boer, E. de (**1985**):
 “Auditory Time Constants: A Paradox?,”
 in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer, Berlin, Heidelberg, New York), 141–158.
- Eddins, D. (**1993**):
 “Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region,”
 J. Acoust. Soc. Am. 93, 470–479.
- Egan, J.P., Lindner, W.A., and McFadden, D. (**1969**):
 “Masking-level differences and the form of the psychometric function,”
 Percept. Psychophys. **6**, 209–215.
- Fassel, R. and Püschel, D. (**1993**):
 “Modulation detection and masking using deterministic and random maskers,”
 in *Contributions to Psychological Acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), 419–429.
- Fassel, R. (**1994**):
Experimente und Simulationsrechnungen zur Wahrnehmung von Amplitudenmodulationen im menschlichen Gehör,
 Doctoral Thesis, University of Göttingen.

- Fassel, R., and Kohlrausch, A. (1995):
“Modulation detection as a function of carrier frequency and level,” IPO
Annual Progress Report **30**, 21–29,
- Fastl, H. (1990):
“The hearing sensation roughness and neuronal responses to AM-tones,”
Hear. Res. **46**, 293–295.
- Fischer, F.A. (1967):
Einführung in die statistische Übertragungstheorie
BI Hochschultaschenbücher, Mannheim, Zürich.
- Fleischer, H. (1981):
“Amplitudenmodulation von Terzrauschen: Experimente und theoretische
Ergebnisse,”
Acustica **47**, 155–163.
- Fleischer, H. (1982):
“Modulationsschwellen von Schmalbandrauschen,”
Acustica **51**, 154–161.
- Fleischer, H. (1982b):
“Calculating psychoacoustic parameters of amplitude modulated narrow
noise bands,”
Biol. Cybern. **44**, 177–184.
- Formby, C. and Muir, K. (1988):
“Modulation and gap detection for broadband and filtered noise signals,”
J. Acoust. Soc. Am. **84**, 545–550.
- Forrest, T.G. and Green, D.M. (1987):
“Detection of partially filled gaps in noise and the temporal modulation
transfer function,”
J. Acoust. Soc. Am. **82**, 1933–1943.
- Glasberg, B.R. and Moore, B.C.J. (1990):
“Derivation of auditory filter shapes from notched-noise data,”
Hear. Res. **47**, 103–138.
- Green, D.M. (1960):
“Auditory detection of a noise signal,”
J. Acoust. Soc. Am. **32**, 121–131.
- Green, D.M. (1967):
“Additivity of masking,”
J. Acoust. Soc. Am. **41**, 1517–1525.

- Green, D.M. (**1985**):
 “Temporal factors in psychoacoustics,”
 in *Time Resolution in Auditory Systems*, edited by A. Michelsen
 (Springer, Berlin, Heidelberg, New York), 122-140.
- Green, D.M. und Dai, H. (**1991**):
 “Probability of being correct with 1 of M orthogonal signals,”
Perception & Psychophysics 49, 100–102.
- Green, D.M. and Swets, J.A. (**1966**):
Signal Detection Theory and Psychophysics
 (Wiley, New York).
- Green, D.M., Berg, B.G., Dai, H., Eddins, D.A., Onsan, Z., and Nguyen, Q. (**1992**):
 “Spectral shape discrimination of narrow-band sounds,”
J. Acoust. Soc. Am. 92, 2586–2597.
- Green, G.G. and Kay, R.H. (**1973**):
 “The adequate stimuli for channels in the human auditory pathways concerned with the modulation present in frequency-modulated tones,”
J. Physiol. (London) 234, 50–52.
- Green, G.G. and Kay, R.H. (**1974**):
 “Channels in the human auditory system concerned with the waveform of modulation present in amplitude- and frequency-modulated tones,”
J. Physiol. (London) 241, 29–30.
- Hall, J.W., III (**1987**):
 “Experiments on comodulation masking release,” in *Auditory Processing of Complex Sounds*, edited by W.A. Yost and C.S. Watson
 (Lawrence Erlbaum Associates, Hillsdale, NJ).
- Hartmann, W.M. and Pumplin, J. (**1988**):
 “Noise power fluctuation and the masking of sine signals,”
J. Acoust. Soc. Am. 83, 2277–2289.
- Holube, I., Colburn, H.S., Par, S. v. d. and Kohlrausch, A. (**1995**):
 “Model simulations of masked thresholds for tones in dichotic noise maskers,”
J. Acoust. Soc. Am. 97, 3411–3412.
- Holube, I., Colburn, H.S., Par, S. v. d. and Kohlrausch, A. (**1995**):
 “Simulationen der Mithörschwellen von Testtönen in dichotischen Rauschmaskierern,”
Fortschritte der Akustik, DAGA '95, 783–786.

- Houtgast, T. (1989):
“Frequency selectivity in amplitude-modulation detection,”
J. Acoust. Soc. Am. 85, 1676–1680.
- Kay, R.H. (1982):
“Hearing of modulation in sounds,”
Phys. Rev. **62**, 894–975.
- Kay, R.H. and Matthews, D.R. (1972):
“On the existence in human auditory pathways of channel selectivity to the modulation present in frequency modulated tones,”
J. Physiol. **225**, 657–667.
- Kohlrausch, A., Püschel, D. and Alpehi, H. (1992):
“Temporal resolution and modulation analysis in models of the auditory system,” in *The Auditory Processing of Speech*, edited by M.E.H. Schouten (Mouton de Gruyter, Berlin, New York), 85-98.
- Kohlrausch, A. and Sander, A. (1995):
“Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,”
J. Acoust. Soc. Am. **97**, 1817–1829.
- Langner, G. and Schreiner, C. (1988):
“Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism,”
J. Neurophysiol. 60, 1799–1822.
- Langner, G. (1992):
“Periodicity coding in the auditory system,”
Hear. Res. **60**, 115–142.
- Lawson, J. L. and Uhlenbeck, G. E. (1950):
Threshold Signals, volume 24 of *Radiation Laboratory Series*,
McGraw Hill, New York.
- Levitt, H. (1971):
“Transformed up-down procedures in psychoacoustics,”
J. Acoust. Soc. Am. 49, 467–477.
- Maiwald, D. (1967a):
“Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen,”
Acustica 18, 81–92.

- Maiwald, D. (1967b):
 “Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas,”
Acustica 18, 193–207.
- Martens, J.P. (1982):
 “A new theory for multitone masking,”
J. Acoust. Soc. Am. 72, 397–405.
- Mc Fadden, D. (1987):
 “Comodulation detection differences using noise-band signals,”
J. Acoust. Soc. Am. 81, 1519–1527.
- Moore, B.C.J. and Glasberg, B.R. (1986):
 “The role of frequency selectivity in the perception of loudness, pitch and time,” in *Frequency Selectivity in Hearing*, edited by B.C.M. Moore, (Academic, London), 251–308.
- Moore, B.C.M., Glasberg, B.R., Plack, C.J. and Biswas, A.K. (1988):
 “The shape of the ear’s temporal window,”
J. Acoust. Soc. Am. 83, 1102–1116.
- Münkner, S. (1993):
Modellentwicklung und Messungen zur Wahrnehmung nichtstationärer Signale
 Doctoral Thesis, University of Göttingen.
- Münkner, S. and Püschel, D. (1993):
 “A psychoacoustical model for the perception of non-stationary sounds,” in *Contributions to psychological acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), 121–134.
- Oxenham, A.J. and Moore, B.C.J. (1994):
 “Modeling the additivity of nonsimultaneous masking,”
Hear. Res. 80, 105–118.
- Patterson, R.D. and Moore, B.C.J. (1986):
 “Auditory filters and excitation patterns as representations of frequency resolution,” in *Frequency Selectivity in Hearing*, edited by B.C.J. Moore, (Academic, London), 123–178.
- Patterson, R.D., Nimmo-Smith, I., Holdsworth, J. and Rice, P. (1987):
 “An efficient auditory filterbank based on the gammatone function,” in Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE, December 14-15.

- Penner, M.J. (1978):
“A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts,”
J. Acoust. Soc. Am. 63, 195–201.
- Plomp, R. (1983):
“The role of modulation in hearing,”
in *Hearing – Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman
(Academic, London), 270–275.
- Plomp, R. and Bouman M.A. (1959):
“Relation between hearing threshold and duration for tone pulses,”
J. Acoust. Soc. Am. 31, 749–758.
- Püschel, D. (1988):
Prinzipien der zeitlichen Analyse beim Hören
Doctoral Thesis, University of Göttingen.
- Rodenburg, M. (1972):
Sensitivity of the auditory system to differences in intensity
Doctoral Thesis, Medical Faculty of Rotterdam.
- Rodenburg, M. (1977):
“Investigations of temporal effects with amplitude modulated signals,”
in *Psychophysics and Physiology of Hearing*, edited by E.F. Evans and J.P. Wilson
(Academic, London), 429–437.
- Schacknow, P.N. and Raab, D.H. (1976):
“Noise-intensity discrimination: Effects of bandwidth conditions and mode of masker presentation,”
J. Acoust. Soc. Am. 60, 893–905.
- Schreiner, C. and Urbas, J.V. (1988):
“Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields,”
Hear. Res. 32, 49–65.
- Schroeder, M.R. (1973):
“An integrable model for the basilar membrane,”
J. Acoust. Soc. Am. 53, 429–434.
- Sheft, S. and Yost, W. (1990):
“Temporal integration in amplitude modulation detection,”
J. Acoust. Soc. Am. 88, 796–805.

- Sek, A. and Moore, B.C.J. (1994):
“The critical modulation frequency and its relationship to auditory filtering at low frequencies,”
J. Acoust. Soc. Am. 95, 2606–2615.
- Strube, H.W. (1985):
“A computationally efficient basilar-membrane model,”
Acustica 58, 207–214.
- van Zanten, G.A. (1980):
“Temporal modulation transfer functions for intensity modulated noise bands,”
in *Psychophysical and Behavioral Studies in Hearing*,
edited by G. van den Brink and F.A. Bilten
(Delft U. P., Delft, The Netherlands), 206-209.
- Verhey, J.L. and Dau, T. (1995):
“Simulations of spectral masking with a model incorporating an optimal decision strategy,” in *Contributions to psychological acoustics*, edited by B. Kollmeier
(Universitätsgesellschaft Oldenburg, Oldenburg), 107–120.
- Viemeister, N.F. (1977):
“Temporal factors in audition: a system analysis approach,”
in *Psychophysics and Physiology of Hearing*,
edited by E.F. Evans and J.P. Wilson
(Academic, London), 419-427.
- Viemeister, N.F. (1979):
“Temporal modulation transfer functions based upon modulation thresholds,”
J. Acoust. Soc. Am. 66, 1364–1380.
- Viemeister, N.F. and Wakefield, G.H. (1991):
“Temporal integration and multiple looks,”
J. Acoust. Soc. Am. 90, 858–865.
- Viemeister, N.F. and Plack, C.J. (1993):
“Time analysis,”
in *Human Psychophysics*, edited by W.A. Yost, A.N. Popper and R.R. Fay
(Springer, New York), 116-154.
- Yost, W.A. and Sheft, S. (1989):
“Across critical band processing of amplitude modulated tones,”
J. Acoust. Soc. Am. 85, 848–857.

- Yost, W.A., Sheft, S., and Opie, J. (1989):
“Modulation interference in detection and discrimination of amplitude modulation,”
J. Acoust. Soc. Am. 86, 2138–2147.
- Zhou, B. (1995):
“Auditory filter shapes at high frequencies,”
J. Acoust. Soc. Am. 98, 1935–1942.
- Zwicker, E. (1952):
“Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,”
Acustica 2, 125–133.
- Zwicker, E. (1953):
“Die Veränderung der Modulationsschwellen durch verdeckende Töne und Geräusche,”
Acustica 3, 274–278.
- Zwicker, E. and Fastl, H. (1990):
Psychoacoustics
(Springer Verlag, Berlin, Heidelberg).
- Zwislocki, J.J., Hellman, R.P. and Verillo, R.T. (1962):
“Threshold of audibility for short pulses,”
J. Acoust. Soc. Am. 34, 1648–1652.

Danksagung

Ich möchte Prof. Dr. Dr. Birger Kollmeier für sein stetes Interesse an der Weiterentwicklung des “Perzeptions-Modells”, für sein in mich gesetztes Vertrauen und nicht zuletzt für die hervorragenden Arbeitsbedingungen danken.

Prof. Dr. Volker Mellert danke ich für die Unterstützung während meiner Arbeit und die Annahme des Korreferats.

Dr. Armin Kohlrausch möchte ich sehr danken für die zahlreichen intensiven Diskussionen, sein großes Interesse an “grundlegenden” Fragestellungen und seine enorme Motivationsfähigkeit auch in kritischen Phasen der Arbeit.

Bei den Mitgliedern des Graduiertenkollegs und der Arbeitsgruppe Medizinische Physik möchte ich mich für die sehr angenehme Arbeitsatmosphäre bedanken.

Jesko Verhey danke ich vor allem für die intensiven Auseinandersetzungen mit der Frage, “ob man das denn nicht auch analytisch rechnen könnte”.

Martin Hansen, Dr. Birgitta Gabriel, Dr. Stefan Uppenkamp, Dr. Stefan Launer und Dr. Andres Sander danke ich für viele anregende Diskussionen und eine schöne gemeinsame Zeit in Oldenburg fernab jeglicher Physik.

Dr. Ralf Fassel und Dr. Stefan Münkner danke ich für die gute Zusammenarbeit auf unserem gemeinsamen Forschungsgebiet, die vielen Hilfen in speziellen Programmierfragen und die netten Tage in Eindhoven und Göttingen.

Prof. Dr. Brian Moore und Dr. Andrew Oxenham möchte ich für die interessante und sehr schöne Zeit in Cambridge danken. Ihre gesunde Skepsis und Kritik gegenüber dem “Modell” hat zudem sehr zum Gelingen dieser Arbeit beigetragen.

Danken möchte ich auch noch Joachim Neumann für die den Umständen eigentlich nicht entsprechende entspannte Atmosphäre in den letzten gemeinsamen letzten Tagen vor unserer Abgabe.

Lebenslauf

Ich wurde am 28. März 1965 in Hannover als erstes Kind von Joachim und Marion Dau, geb. Neumann geboren. Von 1971 bis 1975 besuchte ich die Grundschule am Welfenplatz in Hannover und wechselte 1975 auf das Gymnasium Lutherschule in Hannover.

Nach dem Abitur im Juni 1985 nahm ich zum Wintersemester 1985 das Maschinenbau-Studium an der Universität Hannover auf. Im Oktober 1987 legte ich die Vordiplomprüfung im Fach Maschinenbau ab. Zum Wintersemester 1987 wechselte ich die Fachrichtung und begann mit dem Physikstudium an der Georg-August-Universität in Göttingen. Im Oktober 1989 legte ich dort die Vordiplomprüfung im Fach Physik ab. Im April 1991 begann ich am Dritten Physikalischen Institut unter der Anleitung von Prof. Dr. Manfred Schroeder mit der Anfertigung meiner Diplomarbeit mit dem Thema: “Der Optimale Detektor in einem Computermodell zur Simulation von psychoakustischen Experimenten”. Mein Studium schloß ich am 26. Oktober 1992 mit der Diplomprüfung ab.

Von November 1992 bis November 1995 war ich als Stipendiat Mitglied des Graduiertenkollegs “Psychoakustik” an der Carl-von-Ossietzky Universität Oldenburg. Während dieser Zeit beschäftigte ich mich neben der Anfertigung der vorliegenden Dissertation mit der Betreuung von Diplomanden und der Anleitung von Studenten im Rahmen des Psychoakustik-Praktikums an der Universität Oldenburg.

Im Dezember 1995 war ich als wissenschaftlicher Mitarbeiter an der Universität Oldenburg im Projekt “Psychoakustische Modellbildung” tätig, das von der Deutschen Forschungsgemeinschaft gefördert wurde. Seit Januar 1996 bin ich als wissenschaftlicher Angestellter im Sonderforschungsbereich “Neuronale Grundlagen kognitiver Leistungen” an der Universität Oldenburg beschäftigt.

A new modeling approach of the “effective” signal processing in the auditory system was developed which describes effects of spectral and temporal integration in amplitude-modulation detection and masking. Envelope fluctuations within each auditory channel are analyzed with a modulation filterbank. The parameters of the filterbank are the same for all auditory filters and were adjusted to allow the model to account for modulation detection and modulation masking data with narrowband carriers at a high center frequency. In the detection stage, the outputs of all modulation filters from all excited peripheral channels are combined linearly with optimal weights. To integrate information across time, a “multiple-look” strategy is implemented within the detection stage which allows the model to account for long time constants derived from the data on modulation integration without introducing true long-term integration. Model predictions are compared with both own experimental results and with experimental data from the literature. A large variety of psychoacoustical experiments can be well described by the model. This supports the hypothesis that amplitude fluctuations are processed by modulation-frequency-selective channels. The model might also be used in applications such as psychoacoustical experiments with hearing-impaired listeners, speech intelligibility and speech quality predictions.