

**Auditory spectro-temporal processing  
in the envelope-frequency domain:  
Experiments and models**

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

**Stephan Ewert**  
geb. am 05. Juli 1972  
in Oldenburg

Erstreferent: Prof. Dr. Dr. Birger Kollmeier

Korreferent: Prof. Dr. Volker Mellert

Tag der Disputation: 18.12.2002

Does anybody really know the secret  
Or the combination for this life and where they keep it  
It's kinda sad when you don't know the meaning  
But everything happens for a reason  
[Limp Bizkit, *Take a Look Around*, 2000]



# CONTENTS

<b>INTRODUCTION</b>	<b>1</b>
<b>1 CHARACTERIZING FREQUENCY SELECTIVITY FOR ENVELOPE FLUCTUATIONS</b>	<b>5</b>
I. INTRODUCTION	6
II. METHOD	7
A. Procedure and subjects	7
B. Apparatus	8
C. Stimuli	8
III. EXPERIMENTS	9
A. Masked-threshold patterns	9
B. Psychophysical tuning curves	12
C. Band-widening experiment	13
D. Discussion	15
IV. MODEL PREDICTIONS	17
A. Envelope power spectrum model (EPSM)	17
B. Effects of peripheral filtering	24
C. Discussion	26
V. SUMMARY AND CONCLUSIONS	32
<b>2 EXTERNAL AND INTERNAL LIMITATIONS IN AMPLITUDE-MODULATION PROCESSING</b>	<b>35</b>
I. INTRODUCTION	36
II. MODELS OF AM PROCESSING	38
III. AM-DEPTH DISCRIMINATION WITH NOISE AND PURE-TONE CARRIERS	40
A. Method	40
B. Results and discussion	41
IV. AM DETECTION WITH FROZEN- AND RANDOM-NOISE CARRIERS	45
A. Rationale	45
B. Method	45
C. Results and discussion	46
V. AM MASKING AS A FUNCTION OF MASKER VARIABILITY	49

A.	Rationale . . . . .	49
B.	Method . . . . .	49
C.	Results and discussion . . . . .	50
VI.	OVERALL DISCUSSION . . . . .	53
VII.	CONCLUSIONS . . . . .	55
<b>3</b>	<b>SPECTRO-TEMPORAL PROCESSING IN THE ENVELOPE-FREQUENCY DO-</b>	
	<b>MAIN</b>	<b>57</b>
I.	INTRODUCTION . . . . .	58
II.	EXPERIMENT 1: ENVELOPE-FREQUENCY SELECTIVITY USING SINU-	
	SOIDAL CARRIERS . . . . .	61
A.	Method . . . . .	61
B.	Results . . . . .	62
III.	EXPERIMENT 2: THE ROLE OF ENVELOPE BEATS IN MODULATION	
	MASKING . . . . .	65
A.	Rationale . . . . .	65
B.	Method . . . . .	65
C.	Results . . . . .	66
IV.	EXPERIMENT 3: INTERFERENCE OF ENVELOPE AND VENELOPE PRO-	
	CESSING IN TONE-IN-TONE MASKING . . . . .	68
A.	Rationale . . . . .	68
B.	Method . . . . .	69
C.	Results . . . . .	69
V.	DISCUSSION . . . . .	71
A.	Experimental results . . . . .	71
B.	Possible model structures . . . . .	73
VI.	SUMMARY AND CONCLUSIONS . . . . .	76
<b>4</b>	<b>MODULATION MASKING PRODUCED BY COMPLEX TONE MASKERS</b>	<b>79</b>
I.	INTRODUCTION . . . . .	80
II.	METHODS . . . . .	82
A.	Subjects . . . . .	82
B.	Apparatus and stimuli . . . . .	82
C.	Procedure . . . . .	85
III.	RESULTS . . . . .	85
A.	Effect of signal phase in two-tone complex masking . . . . .	85
B.	Effect of signal phase in three-tone-complex masking . . . . .	86
IV.	MODELING THE RESULTS . . . . .	88
A.	General model structure . . . . .	89
B.	Model-specific parameters . . . . .	89
C.	Model predictions . . . . .	90
V.	DISCUSSION . . . . .	90

A.	Relation to the data from Moore <i>et al.</i> (1999) . . . . .	92
B.	Effect of the harmonic relation between signal and masker components . . . . .	92
C.	Role of compression . . . . .	93
D.	Role of nonlinearities after peripheral compression . . . . .	93
E.	Role of internal noise . . . . .	94
VI.	SUMMARY AND CONCLUSIONS . . . . .	95
<b>5</b>	<b>MECHANISMS OF ENVELOPE-FREQUENCY SELECTIVITY</b>	<b>97</b>
I.	INTRODUCTION . . . . .	97
II.	MODULATION MASKING AS A FUNCTION OF THE MASKER WAVEFORM . . . . .	99
A.	Method . . . . .	99
B.	Results and discussion . . . . .	100
III.	MODEL PREDICTIONS . . . . .	103
A.	Models . . . . .	103
B.	Results and discussion . . . . .	104
IV.	SUMMARY AND CONCLUSIONS . . . . .	106
	<b>SUMMARY AND CONCLUSIONS</b>	<b>109</b>
	<b>APPENDIX</b>	<b>113</b>
	<b>REFERENCES</b>	<b>119</b>
	<b>CURRICULUM VITAE</b>	<b>127</b>
	<b>DANKSAGUNG</b>	<b>129</b>





# INTRODUCTION

Hearing is an important part of the human sensory system which provides us with information about our day-life surroundings and plays a fundamental role in the communication between human beings. The auditory system, like all sensory systems, can be thought of as a chain of processing stages that is meant to transform a physical event into an internal (neural) representation carrying a variety of important information further exploited by cognitive processes. Psychoacoustics tries to establish a functional relationship between the physical properties of a sound incident to the ears and the auditory sensation related to the sound. Models of auditory processing are an extremely helpful tool, offering meaningful interpretations of behaviourally measured data on the one hand, and helping to ask important new questions for psychoacoustic experiments on the other hand. Most natural sounds in our surroundings, including running speech, have a non-stationary characteristics. They show typical temporal fluctuations in their envelope. Especially for sounds which convey information, such as speech and music, much of the information appears to be carried by the changes of the temporal envelope, rather than by the stationary parts (Plomp, 1988; Drullman et al., 1994; Greenberg and Arai, 1998). This seems plausible since even in typical environments (such as rooms) sounds can undergo strong distortions that maintain the basic temporal structure while other aspects (such as the spectral content) might be strongly altered without changing the perceived information contents of the sound. With increasing fluctuation rates, the perception of envelope fluctuations ranges from a variation in the loudness to a temporal structure or rhythm, followed by the impression of roughness and pitch.

Throughout the last years the spectral decomposition of the envelope has been proven to be a powerful concept to describe various phenomena in envelope processing and amplitude modulation detection and masking. Several researchers have described frequency-selective processing in the envelope-frequency domain (e.g., Kay, 1972; Martens, 1982; Bacon and Grantham, 1989; Houtgast, 1989), suggesting that the auditory system performs a spectral decomposition of the envelope, similar to the well-known and established concept of critical bands (Fletcher, 1940) in the audio-frequency domain. Dau *et al.* (1997a,b) presented quantitative predictions with an auditory-processing model that includes a modulation filterbank. This model was shown to account for a variety of detection and masking data in both the audio-frequency and envelope-frequency domains (e.g. Dau *et al.* 1997a,b;

Verhey *et al.*, 1999; Derleth and Dau, 2000; Derleth *et al.*, 2001). However, the filterbank parameters have been determined in experiments which only represent an indirect measure of frequency selectivity. Furthermore, these and other “key” parameters of modulation processing and perception were derived in the framework of the complex model implying that, for example, the shape of the resulting modulation filters might have been influenced by the transfer characteristics of other processing stages within the model. Also, more generally, there still has been some controversy about the existence of such modulation filters, and alternative concepts of periodicity analysis have been suggested.

The main goal of the present thesis is to obtain a deeper understanding of the basic mechanisms and coding principles underlying the processing and perception of envelope fluctuations. The experimental paradigms presented in the different chapters address specific aspects of envelope processing that allow for a step-by-step characterization of these mechanisms. Basic studies of the processing of spectral components in the audio-frequency domain are often associated with questions related to the frequency selectivity of the peripheral auditory system, intensity coding, and combined spectro-temporal effects when more complex sounds are analysed. In analogy, the current thesis attempts to characterize the processing of the envelope of sounds in terms of i) envelope-frequency selectivity, ii) the coding of amplitude-modulation depth, and iii) the processing of complex modulations. Quantitative model predictions are obtained in the different chapters. It is shown that modulation masking, detection and discrimination, as well as the processing of complex envelope waveforms can be successfully accounted for by a relatively simple functional model that uses only a very small number of parameters. While such an approach might not be appropriate for an “ad hoc” expansion to a more general model of auditory processing, it is appealing in its small number of parameters and the transparency of its predictions. The questions addressed in this thesis are not only important for the basic understanding of envelope processing, but may also be of potential relevance for technical applications, such as robust speech recognition, objective sound quality assessment or psychophysically motivated audio codecs.

In Chapter 1 of the thesis, three detection experiments are performed to directly derive the bandwidth and the shape of auditory modulation filters in terms of a critical band concept in the envelope-frequency domain. The experimental paradigms are adopted from classical studies in the audio-frequency domain and transposed to the envelope-frequency domain. The data are compared to predictions of the envelope power spectrum model (EPSM) which is defined in this chapter. The parameters of bandpass modulation filters are adjusted to account for the various modulation detection and masking data.

Chapter 2 addresses the role of internal and external limitations in amplitude-modulation (AM) detection and discrimination. The effects of external noise (as introduced by stochastic stimuli such as random-noise carriers) can be separated from effects of internal-noise processes that limit resolution in conditions with deterministic stimuli (such as pure-tone or frozen-noise carriers). AM-depth discrimination and AM detection with random and deterministic carriers are investigated. In addition, AM masking is in-

investigated in conditions where the degree of variability of a noise masker is changed, such that conditions ranging from entirely frozen to entirely random are tested. The experimental data are described in terms of the envelope power spectrum model from Chapter 1 and compared to predictions obtained with the signal processing model by Dau *et al.* (1997a,b).

In Chapter 3, modulation masking is investigated with different stimulus types for the signal and the masker in order to examine whether beats between modulations play the same role in tone-in-noise, noise-in-tone and tone-in-tone masking as they do in corresponding masking experiments in the audio-frequency domain. The new concept of the “venelope” is introduced which is defined as the envelope of the (ac-coupled) envelope of the stimuli, and is useful to describe the effect of beats between concurrent envelope components of an arbitrary waveform. The chapter also addresses the question to what extent envelope and venelope fluctuations interact with each other when they lie in a similar range of frequencies. Based on the experimental data, a general model structure for the processing of envelope and venelope fluctuations is proposed. Chapter 4 further investigates the processes underlying the extraction of the venelope in the auditory system. Venelope-phase sensitive experiments are conducted in normal-hearing and sensorineural hearing-impaired listeners. The experiments are motivated by an earlier study by Moore *et al.* (1999) where the non-linear processing in the healthy cochlea was suggested as the possible origin for a demodulation of envelope beats to the envelope-frequency domain. Since cochlear damage is assumed to result in a reduction or loss of the compressive nonlinearity in the input-output function of the basilar membrane, the experiment is a critical test for the role of the cochlear nonlinearity in complex envelope processing.

Finally, Chapter 5 tries to clarify the nature of frequency selectivity in envelope processing. The remaining question is whether the frequency selectivity observed in all the different modulation detection and masking experiments reflects a “true” limited-resolution spectral decomposition of the stimulus envelope, or whether it simply reflects a selective process that is tuned to different repetition rates in the envelope. This question is addressed by measuring detection thresholds for a sinusoidal signal modulation in the presence of a squarewave or a sinusoidal masker modulation, both having the same rate of repetition but different content in the envelope spectrum.

The final conclusion points out implications for future, more complex and more general models of auditory signal processing that can be derived from the results of each chapter.



# CHAPTER 1

# CHARACTERIZING FREQUENCY SELECTIVITY FOR ENVELOPE FLUCTUATIONS<sup>a</sup>

## ABSTRACT

Three experimental paradigms were used to specify the auditory system's frequency selectivity for amplitude modulation (AM). In the first experiment, masked-threshold patterns were obtained for signal-modulation frequencies of 4, 16, 64 and 256 Hz in the presence of a half-octave-wide modulation masker, both applied to the same noise carrier with a bandwidth ranging from 1 to 4 kHz. In the second experiment, psychophysical tuning curves (PTCs) were obtained for signal-modulation frequencies of 16 and 64 Hz imposed on a noise carrier as in the first experiment. In the third experiment, masked thresholds for signal-modulation frequencies of 8, 16, 32 and 64 Hz were obtained according to the "classical" band-widening paradigm, where the bandwidth of the modulation masker ranged from 1/8 to 4 octaves, geometrically centered on the signal frequency. The first two experiments allowed a direct derivation of the shape of the modulation filters while the latter paradigm only provided an indirect estimate of the filter bandwidth. Thresholds from the experiments were predicted on the basis of an envelope power-spectrum model (EPSM) which integrates the envelope power of the modulation masker in the passband of a modulation filter tuned to the signal-modulation frequency. The Q-value of second-order bandpass modulation filters was fitted to the masking data using a least-squares algorithm. Q-values of about 1 for frequencies up to 64 Hz suggest an even weaker selectivity for modulation than assumed in earlier studies. The same model also accounted reasonably well for the shape of the temporal modulation transfer function (TMTF) obtained for carrier bandwidths in the range from 1 to 6000 Hz. Peripheral filtering and effects of peripheral

---

<sup>a</sup>This chapter is published as S. D. Ewert and T. Dau, (2000), "Characterizing frequency selectivity for envelope fluctuations", *J. Acoust. Soc. Am.* **108**, 1181-1196.

compression were also investigated using a multi-channel version of the model. Waveform compression did not influence the simulated results. Peripheral bandpass filtering only influenced thresholds for high modulation frequencies when signal information was strongly attenuated by the transfer function of the peripheral filters.

## I. INTRODUCTION

Effects of frequency selectivity in the audio-frequency domain are well established and associated with the existence of independent frequency channels, or critical bands (Fletcher, 1940; Zwicker *et al.*, 1957). Several researchers have suggested a similar spectral decomposition in the envelope-frequency domain (Kay, 1972; Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.*, 1997a,b). Houtgast (1989) performed modulation-masking experiments where the subject's task was to detect a sinusoidal signal modulation in the presence of a fixed narrowband masker modulation, both applied to a broadband noise carrier. He found that for each half-octave-wide masker, centered at 4, 8 and 16 Hz, respectively, the masking pattern showed a peak at the masker frequency. In addition, Houtgast (1989) performed an experiment adopted from the classical band-widening paradigm, as described by Fletcher (1940). The modulation masker was a noise of variable bandwidth centered at the signal frequency of 8 Hz. The spectral density of the masker was held constant in all conditions. Signal modulation thresholds increased up to a masker bandwidth of about 1/2 octave and remained constant for larger bandwidths. Bacon and Grantham (1989) measured detection thresholds for sinusoidal signal modulation in the presence of a sinusoidal masker modulation, both imposed on the same broadband noise carrier. Most modulation masking generally occurred when the signal frequency was at the masker frequency, resulting in bandpass-shaped masking patterns. The results were interpreted as indicating selectivity in the envelope-frequency domain, analogous to the frequency selectivity in the audio-frequency domain. No quantitative derivations of the bandwidth and shape of the hypothetical filter mechanism were undertaken in these studies. Furthermore, the above experiments were restricted to relatively low signal modulation frequencies, up to 64 Hz.

In a more recent study, Dau *et al.* (1997a,b) examined AM detection thresholds using narrowband noise carriers (see also Fleischer, 1982). The noise carrier bandwidths were 3, 31 and 314 Hz, respectively. The rate of the random envelope fluctuations inherent in the noise carrier increased with increasing bandwidth, resulting in a very different shape of the temporal modulation transfer function (TMTF) for these conditions. These findings could be described qualitatively by assuming that detection of the signal modulation was hardest when its frequency fell within the spectral range of the inherent envelope fluctuations of the carrier. Dau *et al.* (1997a,b) explained the results quantitatively in terms of a model based on a modulation filterbank. Further evidence for such a mechanism was demonstrated in a more recent study by Dau *et al.* (1999). TMTFs were obtained with different types of narrowband noise as the carrier, having the same overall level and a bandwidth of 50 Hz.

The three carriers, Gaussian noise, multiplied noise and low-noise noise produced very different TMTFs for modulation rates from 3 to 100 Hz. The results could be explained reasonably well by taking into account the different shapes of the carrier envelope spectra and by assuming that the amount of inherent envelope power of the carrier that passes a specific modulation filter determines detection threshold.

Thus, the modulation filterbank concept appears to be a powerful tool for describing TMTFs in various stimulus configurations. However, the filterbank parameters have been determined in experiments which only represent an indirect measure of the frequency selectivity. Furthermore, the parameters were originally derived in the framework of a relatively complex signal processing model (Dau *et al.*, 1996a,b, 1997a,b). Thus, the shape of the resulting modulation filters was probably influenced by the transfer characteristic of the adaptation stage prior to the modulation filterbank. Although the adaptation stage represents a major processing stage within the framework of the whole processing model, the role of its time-dependent compressive properties for the processing of modulations is still unclear.

The first part of the current article presents new experimental data which reflect more direct measures of the modulation filter shape. All experimental paradigms were adopted from classical studies performed in the audio-frequency domain: Masked-threshold patterns, psychophysical tuning curves, and masking patterns according to the band-widening paradigm. Data were obtained for a large range of signal frequencies (up to 256 Hz) to examine the “existence region” of modulation filters.

In the second part of the study, simulations are presented which are based on a model referred to here as the “envelope power spectrum model” (EPSM) of modulation masking. The Q-value of the modulation filters assumed in this model is fitted to the experimental masking data. The same model parameters are used for the simulation of experimental TMTF data from the literature. The question is whether the shape and amount of modulation masking as well as the shape of the TMTF and the overall sensitivity to AM can be explained by the same underlying mechanisms. Furthermore, the influence of peripheral filtering and peripheral (waveform) compression are investigated in the framework of a multi-channel version of the EPSM. Capabilities and limitations of the model are discussed.

## II. METHOD

### A. Procedure and subjects

Modulation detection thresholds were obtained using an adaptive three-interval forced-choice (3IFC) procedure. The stimuli were presented in three consecutive intervals separated by 500-ms silent intervals. Two of the three intervals contained the reference stimulus. Depending on the specific experiment, this reference stimulus was either an unmodulated noise carrier (TMTF condition) or a noise on which a stochastic masker modulation was imposed (masking condition). In one randomly chosen interval, the si-

sinusoidal signal modulation was imposed on the same carrier. The subject’s task was to identify the interval containing the signal modulation. During a threshold run, the modulation index  $m$ , in dB ( $20 \log m$ ), was adjusted according to a two-down one-up rule (Levitt, 1971) which provides an estimate of the modulation index necessary for 70.7 % correct responses. The modulation depth was varied by using a 4 dB step size at the beginning of the adaptive track. After each second reversal the step size was divided by 2 until it reached 1 dB. At this step size, eight reversals were obtained and threshold was calculated as the median value of  $20 \log m$  at these reversals. The subjects received visual feedback after each response. The measurement procedure was repeated at least three times for each configuration and subject. For each subject, the final modulation detection threshold was taken as the average across the last three threshold estimates. Unless otherwise stated, all figures show mean data and standard deviations calculated by averaging across three subjects. All subjects had experience in psychoacoustic experiments and had clinically normal hearing. They were between 24 and 33 years old and were paid for participation in the study.

## B. Apparatus

All acoustic stimuli were digitally generated at a sampling frequency of 32 kHz. The stimuli were converted to analog signals by a two-channel 16-bit D/A converter including reconstruction filtering. The stimuli were presented diotically via headphones (Sennheiser HD 25) in a sound-attenuating booth. Signal generation and presentation during the experiments were computer controlled using a signal-processing software package developed at the Drittes Physikalisches Institut at the University of Göttingen. Analytical calculations and computer simulations were performed with Matlab.

## C. Stimuli

In all experiments, the carrier duration was 600 ms including 50-ms  $\cos^2$  onset and offset ramps. Both the sinusoidal signal modulation and the masker modulation had a duration of 500 ms including 50-ms  $\cos^2$  ramps and were temporally centered in the carrier. In all conditions the signal modulation started at a positive going zero-crossing. Gaussian noise (GN) and low-noise noise (LNN) were used as modulation maskers. GN was generated as a sequence of random samples from a Gaussian distribution and was restricted to the required bandwidth by setting the Fourier coefficients outside the desired passband to zero. LNN was originally described by Pumplin (1985) and Hartmann and Pumplin (1988, 1991). They generated LNN by defining the amplitude of its spectral components, while optimizing the phase of these components using a gradient search procedure to achieve a temporal waveform with minimized envelope (power) fluctuations. This procedure results in a temporal waveform with a smooth envelope. It is, however, quite time consuming and not easily implemented. In the present study, a similar waveform was generated in a way described by Kohlrausch *et al.* (1997): the temporal waveform of a



bandlimited GN was divided by its envelope on a sample-by-sample basis. Spectral splatter introduced by this process was then removed by restricting the stimulus to its original bandwidth. These two steps were iterated 10 times. The resulting waveform provides a smooth envelope and small power fluctuations. Although the generation algorithm differs considerably from the algorithm proposed by Hartmann and Pumplin, the noise waveform will be referred to as LNN in the following.

The rationale for using LNN as masker in the present study was to introduce a stimulus which can also be applied at large energies without producing overmodulation when added to the signal. This should, for instance, allow the measurement of masking patterns with a maximum dynamic range (in the modulation domain). This is interesting since, for broadband noise carriers, the reference detection thresholds (as reflected in the TMTF) increase with increasing signal frequency causing a reduced available dynamic range for observing masking at these higher frequencies.

In the case of the LNN as masker, a long-duration realization of approximately 8 sec ( $2^{18}$  samples) was generated only once in advance and was stored to disk. In the case of the GN as masker, a realization of approximately two seconds ( $2^{16}$  samples) was generated before each threshold run. In each presentation interval during the experimental run, the masker was cut out randomly from the long realization.

Signal and masker were applied consecutively to a wideband noise carrier, resulting in the envelope  $(1+M_m(t))(1+M_s(t))$  where  $M_m(t)$  reflects the masker modulation and  $M_s(t)$  reflects the signal modulation. This multiplicative approach was also used by Houtgast (1989). Unless otherwise stated the carrier was restricted to the 1 to 4 kHz band after applying the modulation. The reduction in effective modulation depth was less than 1 dB for the largest signal modulation frequency applied. An unmodulated and uncorrelated inverse bandstop-filtered noise of the same type was added at the same spectrum level. Thus, the resulting stimuli in the signal and the nonsignal interval(s) always had the same long-term power spectrum. The spectral manipulations were performed in the audio-frequency domain by adjustment of the Fourier coefficients of the 2-s signal realizations.

The average power of a sinusoidally amplitude modulated stimulus is increased by the factor  $1 + m^2/2$  compared to the unmodulated stimulus. For large modulation depths, detection might therefore be based on changes in the overall intensity rather than on the presence or absence of modulation. To eliminate level cues, the waveforms were adjusted to have equal energy in each interval of the forced-choice trial. The overall level of the stimuli was 65 dB SPL, equivalent to a spectrum level of 23 dB.

### III. EXPERIMENTS

#### A. Masked-threshold patterns

Masked-threshold patterns (MTPs) are commonly considered as a very useful paradigm to investigate the auditory system's frequency selectivity in the audio-frequency domain. Generally, the threshold for a sinusoidal signal is obtained in the presence of a narrowband

masker of fixed frequency as a function of the signal frequency. Houtgast (1989) and Bacon and Grantham (1989) adopted this experimental paradigm to the modulation domain. They measured detection thresholds for a sinusoidal amplitude modulation in the presence of an additional masker modulation, applied to a broadband noise carrier. The masker modulation was a narrow-band noise in the Houtgast study and a tone in the study by Bacon and Grantham. The spectral position of the masker modulation was fixed in the two studies. Both studies showed most masking when the signal frequency was at the masker frequency. The peaked patterns obtained for a certain masker frequency may be considered as reflecting the “modulation excitation pattern” evoked by the modulation masker, analogous to the excitation patterns commonly described in the audio-frequency domain (e.g., Zwicker and Feldtkeller, 1967; Moore and Glasberg, 1986). The shape of the masked threshold pattern for a particular masker frequency does not, of course, directly reflect the shape of a specific modulation filter. A more direct approach to characterize the modulation-filter shape at a specific “position” would be to measure the threshold for the (fixed) signal frequency in the presence of the masker frequency whose spectral position is varied relative to the signal frequency. This fixed signal modulation frequency approach was used by Takahashi and Bacon (1992) and Lorenzi *et al.* (1997), and was also used in the present study for deriving MTPs. Such a paradigm has sometimes also been used in the audio-frequency domain (Chistovich, 1957; Rodenburg *et al.*, 1974).

## 1. Conditions

The GN was used as the carrier. The signal frequency was 4, 16, 64 or 256 Hz. For each signal frequency, the spectral position of the masker band was varied in the range from -2 to +2 octaves relative to the signal frequency, using a step size of 2/3 octaves. The bandwidth of the masker was 1.4, 5.6, 22.3 and 89.2 Hz for the signal frequencies of 4, 16, 64 and 256 Hz, respectively. In each of the on-frequency conditions (where the masker was centered geometrically at the signal frequency) this corresponds to a 1/2-octave bandwidth. We have chosen such a configuration because previous studies suggested a logarithmic scaling of the modulation filter bandwidth. In the case of GN as the modulation masker, the rms modulation depth of the masker was -10 dB in all conditions. In the case of LNN as the masker, the rms modulation depth of the masker was -10 dB for the signal frequencies 4 and 16 Hz while it was -7 dB for the signal frequencies of 64 and 256 Hz. Since LNN exhibits less envelope fluctuations than GN the masker level could be increased by 3 dB without producing overmodulation when applied to the signal<sup>1</sup>.

In addition, the TMTF was measured for the same carrier in order to obtain reference thresholds for the signal frequencies without any external masker modulation.

---

<sup>1</sup>This was done in an attempt to obtain about the same level for the masked thresholds at the peak frequencies, for all signal frequencies tested. Preliminary experiments showed that, in case of the LNN masker, this could be obtained with about a 3-dB higher masker level for the signal frequencies 64 and 256 Hz than for the signal frequencies 4 and 16 Hz, without producing overmodulation.

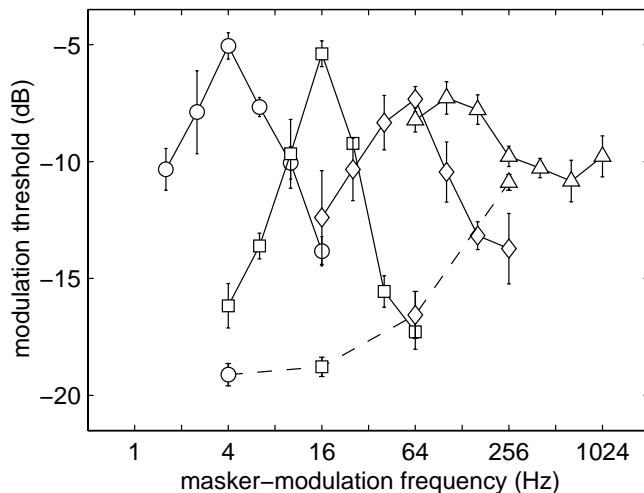


FIG. 1.1: Average masked-threshold patterns (MTPs) for signal frequencies of 4 Hz (circles), 16 Hz (squares), 64 Hz (diamonds) and 256 Hz (triangles). The signal modulation depth at threshold is plotted as a function of the masker frequency. The masker level was always  $-10$  dB. In each on-frequency condition, the GN masker was  $1/2$  octave wide. The absolute bandwidth was held constant when the masker was shifted in the range from  $-2$  to  $+2$  octaves relative to the signal frequency (see text). The dashed curve represents the “reference” thresholds without masker modulation. The carrier was a bandpass GN ranging from 1 to 4 kHz.

## 2. Results

Figure 1.1 shows average masked-threshold patterns, obtained with a carrier noise band in the range from 1 to 4 kHz. Signal modulation depth at threshold is plotted as a function of the masker frequency. In this condition, the masker was a GN. The four peaked patterns represent the data for the signal frequencies of 4 Hz (circles), 16 Hz (squares), 64 Hz (diamonds) and 256 Hz (triangles).

The patterns for 4, 16 and 64 Hz show a maximum masking effect when the masker is centered at the signal frequency. The shape of the pattern is approximately symmetric on a logarithmic frequency scale. The pattern for 256 Hz looks quite different from the pattern at the lower signal frequencies. Masking can only be observed for masker frequencies below the signal frequency. For masker frequencies at and above the signal frequency, hardly any masking effect can be observed; masked thresholds are about the same as the reference threshold for 256 Hz in the TMTF condition. The maximum amount of masking, given by the difference between masked threshold and corresponding reference threshold, is about 14 dB for the signal frequencies 4 and 16 Hz, and 10 and 4 dB for the signal frequencies 64 and 256 Hz.

Figure 1.2 shows corresponding data obtained with LNN instead of GN as the masker. The masking patterns for 4, 16 and 256 Hz are very similar to those obtained with the GN masker. For 64 Hz, the pattern is somewhat broadened compared to the GN condition and has a slightly asymmetric shape. At this frequency the peak of the pattern is shifted by  $-2/3$  octaves towards lower envelope frequencies. The slope of the low-frequency tail of the pattern is somewhat reduced compared to that obtained with GN. For all signal frequencies the peak threshold is at about  $-5$  dB. Note that the patterns for 64 and 265 Hz are shifted by about 3 dB towards higher values compared to the patterns obtained with GN, as a consequence of the higher level for the LNN masker. Again, since the modulation

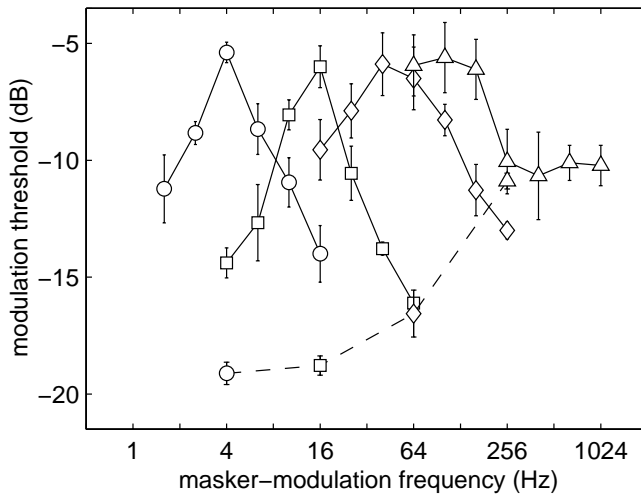


FIG. 1.2: Same as in Fig. 1.1 but with LNN as the masker modulation instead of GN. The carrier was the same as in the previous experiment.

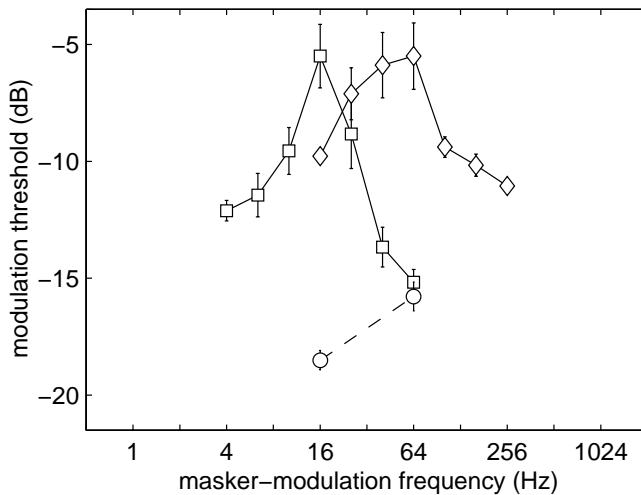


FIG. 1.3: Same as in Fig. 1.2 but with a GN carrier ranging from 4 to 7 kHz instead of 1 to 4 kHz.

thresholds of the TMTF increase with increasing frequency, the resulting masking effect at the peak frequencies decreases with increasing frequency.

Figure 1.3 shows masked-threshold patterns for 16 and 64 Hz obtained with a carrier noise band in the range from 4 to 7 kHz. Thresholds connected by the dashed line represent the TMTF condition (as reference condition without masker modulation). Both masked thresholds as well as reference thresholds are very similar to those obtained in the low-frequency-carrier condition from the previous figure.

## B. Psychophysical tuning curves

This experiment measures psychophysical tuning curves (PTCs) in the envelope-frequency domain, which are also adopted from a popular measure of the frequency selectivity in the audio-frequency domain. The PTC is meant to represent a perceptual correlate of a standard physiological measure, the frequency threshold curve (FTC), commonly referred to as the tuning curve. The FTC is obtained by measuring the level of a sinusoidal stimulus at a certain criterion response in a primary auditory neuron, as a function of the frequency of the stimulus (Kiang *et al.*, 1965). The frequency requiring

the lowest level is referred to as the characteristic frequency (CF) of the auditory neuron. As the stimulus frequency diverges from the CF, the level needed to achieve a constant activity of the neuron usually rises.

To obtain a PTC, the signal is presented to the subject at a fixed level and frequency in the presence of a narrowband masker whose spectral position is varied. The PTC thus reflects the level of the masker at detection threshold for the signal as a function of the masker frequency. It is generally assumed that when the signal level is close to its threshold, it will stimulate only a small group of neurons with similar CF and the PTC paradigm might become analogous to a corresponding physiological tuning curve.

## 1. Conditions

The PTCs in the modulation domain were obtained for 16 and 64 Hz. The level of the signal modulation was kept constant at 4 dB above the average detection threshold ( $-15$  dB for 16 Hz and  $-13$  dB for 64 Hz). At this modulation depth the signal modulation was clearly detectable for all subjects. As for the MTPs from the previous experiment, the masker was a 1/2-octave-wide LNN (relative to the signal frequency) which was presented in the range from 2 octaves below to 2 octaves above the signal frequency. GN was not tested. The subject's task was to adjust the masker level so that it just masked the signal. The starting level of the masker was 5 dB below the signal level (near threshold)<sup>2</sup>.

## 2. Results

Figure 1.4 shows average PTCs. Masker modulation depth at signal threshold is plotted as a function of the masker frequency. The shape of the PTCs largely corresponds to the shape of the corresponding MTPs from Fig. 1.2. For 64 Hz, the peak of the PTC is shifted by one data point towards lower frequencies as it was also observed for the MTP. The dynamic range of the PTC is 3 dB larger (10 dB) than that of the MTP (7 dB). For 16 Hz, the PTC is tuned slightly broader than the corresponding MTP. For both signal frequencies (16 and 64 Hz), the two paradigms do not show differences in the symmetry of the patterns.

## C. Band-widening experiment

The band-widening experiment represents a classical experimental approach to characterizing the bandwidth of a peripheral filter and was first described by Fletcher (1940). The basic idea behind this paradigm is to determine the filter bandwidth from the shape of the threshold curve obtained with a pure tone presented in noise with variable bandwidth. The noise is generally centered at the signal frequency. Since the spectral power density

---

<sup>2</sup>If GN were used as the masker instead of LNN, in this experiment, masker levels larger than  $-10$  dB would have led to overmodulation when combined with the signal. A slightly larger dynamic range in case of the GN masker may have been obtained with a decreased signal level. However, the signal level was already only 4 dB above threshold.

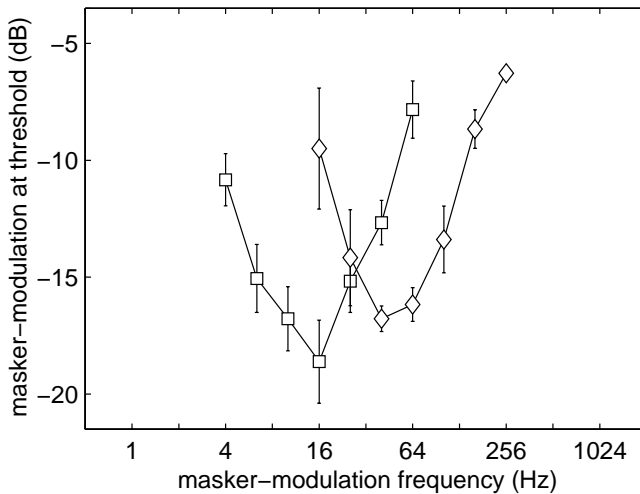


FIG. 1.4: Average psychophysical tuning curves (PTCs) for the signal frequencies 16 Hz (squares) and 64 Hz (diamonds). The level of the LNN masker at signal threshold is plotted as a function of the masker frequency. The signal was presented at a fixed level of 4 dB above the average detection threshold. Masker bandwidth and center frequency were the same as in the corresponding MTP experiment from Fig. 1.2.

is held constant, the total power of the masker increases as the bandwidth increases. Assuming auditory filters to be rectangular in shape as a first approximation, and assuming thresholds to be proportional to the noise power at the output of the filter, the following shape of the threshold curve should be expected: As long as the noise bandwidth is smaller than the filter bandwidth threshold will increase by 3 dB per doubling of the noise bandwidth, while threshold should remain constant for larger bandwidths. The threshold curve could then be fitted by two lines and the bandwidth corresponding to the point of intersection provides an estimate of the “critical bandwidth” (Fletcher, 1940; Bos and deBoer, 1966).

In the present study, as in the study by Houtgast (1989), the band-widening technique was transposed to the modulation domain to estimate the width of the modulation channels. In fact, modulation filters (as well as peripheral filters) are far from rectangular in shape. Nevertheless, as long as the filter function exhibits a relatively sharp roll-off on both sides of its passband, the threshold function obtained in the band-widening experiment should show a narrow transition between the increasing and constant part of the function.

## 1. Conditions

The band-widening experiment was performed for signal frequencies of 8, 16, 32 and 64 Hz. A LNN masker was used which was centered geometrically at the signal frequency. The bandwidth of the masker ranged from 1/2 to 4 octaves. The signal and masker were imposed on a broadband GN carrier. For a signal frequency of 8 Hz, the experiment was repeated with a GN masker and a pink-noise carrier to most closely reproduce the experimental conditions described by Houtgast (1989). As in the Houtgast study, the stimuli were restricted to 1-4 kHz after modulation. In addition, the inverse bandstop-filtered carrier noise was added at the same spectrum level.

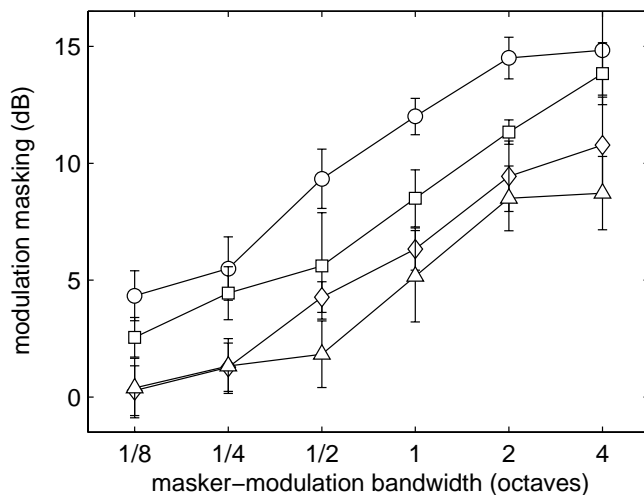


FIG. 1.5: Average data obtained in the band-widening experiment. Masking is plotted as a function of the masker bandwidth for the signal frequencies 8 Hz (circles), 16 Hz (squares), 32 Hz (diamonds), and 64 Hz (triangles). Masking was derived by subtracting the reference threshold from the masked threshold. The masker was a LNN which was geometrically centered at the signal frequency. As in most of the previous conditions, the carrier was a GN ranging from 1 to 4 kHz.

## 2. Results

Figure 1.5 shows the data obtained with LNN as the masker. All data points represent the difference in thresholds between the masked condition and the reference condition (TMTF), expressed as masking (in dB). Masking is plotted for the signal frequencies 8 Hz (circles), 16 Hz (squares), 32 Hz (diamonds) and 64 Hz (triangles), respectively. For each signal frequency, threshold increases with increasing masker bandwidth. This increase is observed for the whole range of masker bandwidths. For 8 and 64 Hz, thresholds appear to asymptote at the largest masker bandwidths (2 and 4 octaves). For each bandwidth masking decreases with increasing signal frequency. For the largest bandwidth condition, masking amounts to 15 dB for 8 Hz and only 8 dB for 64 Hz.

Figure 1.6 shows the data for 8 Hz obtained with a GN masker as in the Houtgast study. For direct comparison, the data from Houtgast are indicated by the dashed curve. In the data of Houtgast, thresholds first increase at a rate of 3-4 dB per octave increase of the noise bandwidth, and then remain constant for bandwidths larger than 1/2 octave. In contrast, our own data (indicated by the solid curve) do not show such a behavior. Thresholds show an almost continuous increase with increasing bandwidth, as was also observed for the LNN masker.

## D. Discussion

The data obtained in the present study provide further psychoacoustical evidence for frequency-selective mechanisms involved in the processing of amplitude modulation. The MTPs showed a basically symmetric shape on a logarithmic frequency scale for signal frequencies up to 64 Hz. The shape and bandwidth of the patterns correspond very well with those of the PTCs, suggesting that the principal underlying mechanism is the same. No tuning was observed for the signal frequency of 256 Hz. Probably, two different processes are involved, one responsible for modulation tuning and a second one responsible for the general decrease of sensitivity, or sluggishness, at frequencies above about 100-

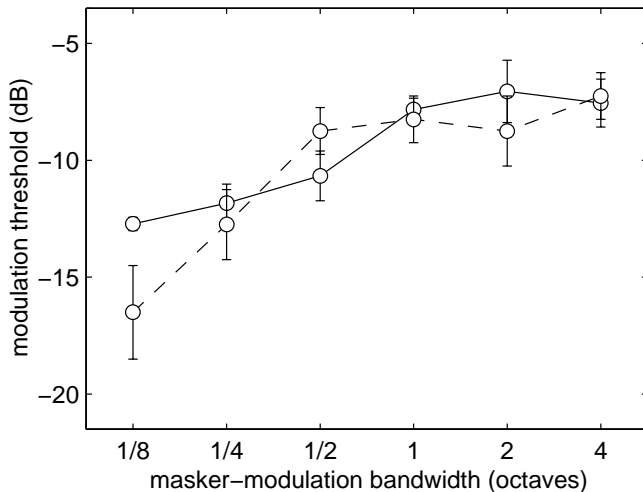


FIG. 1.6: Comparison of the data from the present study and the data of Houtgast (1989). The masked threshold for 8 Hz is plotted as a function of the masker bandwidth. The masker was GN instead of LNN as in the previous experiment. The solid curve represents our own data while the dashed curve is redrawn from Houtgast (1989). The experimental conditions were the same in both studies.

150 Hz. This will be further investigated in the modeling section of the present article.

Differences were observed between the MTPs obtained with LNN and GN as the modulation masker. The patterns appeared to be slightly more sharply tuned for GN than for LNN. Although the use of LNN has some appeal because of its low-peak factor, it also may complicate the interpretation of the data, since additional detection cues may become available. As we realized during the experiments, modulation beatings between the signal frequency and the instantaneous frequency of the masker can occur and influence the shape of the masking pattern. Thresholds appeared to be most influenced in the on-frequency condition where the perceptual salience of the beats was strongest. This is also in line with recent data by Moore *et al.* (1999), who showed that the perceptual salience of modulation beating at low rates is strong enough to influence the detection threshold for signal modulation at a similar rate. Thus, the resulting pattern obtained with LNN as a modulation masker may underestimate the amount of frequency selectivity for modulation. Using a GN as the masker, modulation beatings probably play a minor role since they are at least partly masked by the intrinsic fluctuations of the masker.

The experimental data obtained in the band-widening experiment differ from those published by Houtgast (1989). While the general form of the data in the Houtgast study was quite similar to that generally obtained in corresponding experiments in the spectral domain, the current study only showed a general increase of masking with increasing masker bandwidth. In fact, one might argue that only the data point at a masker bandwidth of 1/8 octave clearly differs in both studies while all other data points lie within 2 dB (Fig. 1.6). However, this exactly reflects the problem behind the band-widening experiment: differences in single data points can lead to very different estimates of the underlying selectivity which makes this paradigm unreliable. Our own data do not allow any direct estimation of the underlying modulation filter characteristics. Qualitatively, the filters appear to broaden with increasing center frequency since the amount of masking decreases with increasing signal frequency. The fact that the slope of the curves is similar for all signal frequencies as a function of the masker bandwidth indicates that the filter



bandwidth does not change much on a *logarithmic* frequency scale. These observations are consistent with the observations from the previous experiments. Thus, the data obtained in the band-widening experiment may still be interpreted in favor of modulation channels. However, these data alone, of course, would hardly allow a reliable estimation of absolute bandwidth and shape of the filter, a problem which also has been described in a similar way for the audio-frequency domain (Patterson and Moore, 1986).

In order to derive quantitative values for the modulation filter parameters and to test whether the experimental data of the study can be described consistently, model predictions are needed. These are presented in the following section.

## IV. MODEL PREDICTIONS

### A. Envelope power spectrum model (EPSM)

In this section, the data obtained in the present study are compared with model predictions. The model contains three basic stages: envelope extraction of the stimuli, bandpass filtering in the envelope-frequency domain, and calculation of the ac-coupled power at the output of a specific (modulation) bandpass filter. Since the principal structure of the model is very similar to the structure of the well-known power-spectrum model of masking (Fletcher 1940; Patterson and Moore, 1986) but transposed to the envelope-frequency domain, it will be referred to as the “envelope power spectrum model” (EPSM) in the following. Within the EPSM it is assumed that AM-detection thresholds are related to the total envelope noise power that passes through the modulation filter centered at the signal frequency.

This modeling approach is equivalent to the calculations presented in a recent study by Dau *et al.*(1999), where TMTFs were predicted for a large set of carrier bandwidths and for different types of noise as the carrier. In contrast to the calculations in Dau *et al.*(1999), the present study generally deals with modulation *masking* conditions where the non-signal interval contains a carrier which is already modulated. Thus, in these conditions the envelope power spectrum of the non-signal stimulus is given by the convolution of the envelope power spectrum of the carrier alone with the envelope power spectrum of the modulator (see the Appendix A).

To calculate the envelope power of modulated noise falling in the transfer range of a specific modulation filter (centered at a frequency  $cf_{env}$ ), the envelope power spectrum of the modulated noise,  $N_{(m)}$ , must be multiplied by the squared transfer function of the modulation filter,  $W_{cf}(f_{env})$ , and then integrated across envelope frequency  $f_{env}$ . The ac-coupled envelope power,  $P_{env}(cf_{env})$ , can then be written as

$$P_{env}(cf_{env}) = \frac{1}{N_{(m)}(0)} \int_{f_{env}>0}^{\infty} N_{(m)}(f_{env}) W_{cf}(f_{env}) df_{env}. \quad (1.1)$$

Here  $N_{(m)}(0)$  denotes the dc-value of the envelope power. If the carrier is an unmodulated bandlimited GN, its envelope power spectrum,  $N$ , can be calculated according to the

formula by Lawson and Uhlenbeck (1950). They showed that, assuming a rectangular shape of the power spectrum of the Gaussian noise carrier, its envelope power spectrum approximately has a triangular shape besides the dc peak and stretches up to the frequency that equals the carrier bandwidth (for details, see also Dau *et al.*, 1999).

One way to relate the calculated envelope power to an experimentally obtained AM-detection threshold is to assume a certain signal-plus-noise-to-noise  $(S + N)/N$  ratio at threshold. In the calculations presented here, a 1-dB  $(S + N)/N$  criterion was used, assuming that a 1-dB increase in the envelope power at the output of the modulation filter is necessary to detect the signal modulation. For computational reasons, the  $(S + N)/N$  criterion was converted into a  $S/N$  ratio at threshold. Under the assumption that signal and noise are statistically independent (which is the case in the present study) the corresponding signal-to-noise ratio is  $S/N \approx -6$  dB. In order to introduce a lower limit for the integrated envelope power, a value corresponding to an AM-detection threshold of  $-30$  dB is finally added after integration (see Dau *et al.*, 1999). This value was intended to reflect the empirically found “absolute” detection threshold for a signal modulation imposed on a sinusoidal carrier at a medium level (e.g., Viemeister, 1979; Kohlrausch *et al.*, 2000)<sup>3</sup>.

The present model does not incorporate the influences of spectral filtering after modulation. Filtering after modulation reduces the effective modulation depth of the modulated stimulus — something that is not included in the calculations. Furthermore, the inverse bandstop filtered masker noise, which was added in the experiments, was not considered in the calculations: the auditory system is assumed to be “focused” at the region between lower and upper spectral edge of the stimuli. More importantly, the current version of the EPSM does not include peripheral filtering, adaptation or any influence of the statistics of the integrated envelope power which is a random variable for short-term intervals. In this respect, the EPSM model differs substantially from the more complex processing model described in Dau *et al.* (1997a,b) despite the similar modulation filterbank approach. In particular, the complex model contains an adaptation stage which allows the description of *both* simultaneous and nonsimultaneous masking (Dau *et al.*, 1996a,b; Derleth and Dau, 2000) as well as modulation detection data (Dau *et al.*, 1997a,b). Such a stage is not contained within the EPSM. In addition, the detection stage within the complex model performs a kind of pattern recognition of the whole temporal course of the internal representation of the stimuli. Decisions are based on the cross correlation between the internal representation of the actual stimulus and a normalized suprathreshold template, and an adaptive tracking procedure is used as in the real experiment. The EPSM is a model which is restricted to amplitude modulation processing and makes only a few simple assumptions about the processing of modulated stimuli. Nevertheless, even if this model probably oversimplifies the “real” processing in the auditory system (more than does the complex model) this analytical approach might be helpful for the understanding

---

<sup>3</sup>Sinusoidal carriers do not exhibit inherent fluctuations like noise carriers. Thus modulation detection must be limited by *internal* noise.

TABLE 1.1: Best fitting Q-values for the second-order bandpass filters assumed within the framework of the EPSM. The Q-value was fitted to the masking patterns derived from the Figs.1.1-1.3. A least-square fitting procedure was used. Corresponding  $-3$ -dB bandwidths of the filters are given in parentheses. The asterisks indicate that a Q-value could not be derived from the data. The bars denote conditions where no experimental data were obtained.

$cf$ (Hz)	type of masker noise, carrier range		
	GN, 1-4 kHz	LNN, 1-4 kHz	LNN, 4-7 kHz
4	0.718 (5.6)	0.711 (5.6)	-
16	1.713 (9.3)	1.255 (12.7)	0.981 (16.3)
64	0.728 (87.9)	0.411 (155.7)	0.315 (203.2)
256	*	*	-

of what processing stages are essential for successfully describing modulation detection and masking phenomena.

## 1. Fitting modulation filters to the masking data

The shape of the experimental MTP, obtained in Section III., was basically symmetrical on a logarithmic frequency scale, for signal frequencies up to 64 Hz. In a previous study it was observed that the general sensitivity for modulation decreases above about 150 Hz (Kohlrausch *et al.*, 2000). Such a behavior was found for sinusoidal carriers at high frequencies where neither intrinsic carrier fluctuations nor spectral cues can influence detection (at least up to several hundred Hertz). Based on these observations, the modulation filterbank proposed here consists of an array of second-order bandpass filters combined with an additional first-order lowpass filter with a cut-off frequency of 150 Hz. With the Q-value as the only free parameter within the model, the best-fitting filter function for each empirical MTP was obtained using a least-squares fitting procedure. The squared deviation between predicted and measured masking was minimized using a Nelder-Mead-type simplex search method.

Table 1.1 shows the best fitting Q-values for the different signal frequencies ( $cf$ ) 4, 16, 64 and 256 Hz for all conditions tested in the experiments, i.e., with GN and LNN as the modulation masker and with the low- and high-frequency noise band as the carrier, respectively. The values in parentheses represent the corresponding  $-3$ -dB bandwidths of the filter.

The asterisk for 256 Hz indicates that it was not possible to derive a Q-value from these data. The derived Q-value for the 4-Hz filter is about the same for GN and LNN while for the signal frequencies 16 and 64 Hz it differs considerably for the two noises, indicating sharper tuning for GN than for LNN. However, as already addressed in Section D., the LNN pattern might have been influenced by cues resulting from modulation beating, particularly in the on-frequency condition, which would result in an underestimation of the system's "real" amount of envelope frequency selectivity.

The upper panel (a) of Fig. 1.7 shows model predictions (closed symbols) of the masking patterns, obtained with the best-fitting values for the modulation filters, together with the corresponding experimental masking data (open symbols, replotted from Fig. 1.1). The predicted patterns agree very well with the experimental data. The middle panel (b) of Fig. 1.7 shows corresponding predictions where, for reasons of simplicity, a constant Q-value of 1 was assumed for all modulation bandpass filters. The correlation between model and data still is quite reasonable, except perhaps for 16 Hz where the predicted masking pattern is too broad.

In earlier studies (Dau *et al.*, 1997a,b, 1999) where the shape of modulation filters was estimated only indirectly, first-order resonance filters were assumed with a constant absolute bandwidth of 5 Hz for modulation frequencies up to 10 Hz and a Q-value of 2 for filters centered at frequencies larger than 10 Hz. The lower panel (c) of Fig. 1.7 shows model predictions where these filters were incorporated in the EPSM. For the frequencies 64 and 256 Hz, the patterns show too sharp tuning while for 4 and 16 Hz the low-frequency skirt is too shallow. For direct comparison, Fig. 1.8 shows the transfer functions of the filters (dashed lines) suggested in the earlier studies in the framework of the more complex signal processing model together with those for the band-pass filters with a Q-value of 1 (solid lines), as suggested in the present study.

Finally, Fig. 1.9 shows model predictions from the band-widening experiment described in Section III. In this and in the following simulations, a constant Q of 1 was assumed for all signal frequencies. First, as observed in the experimental data from Fig. 1.5, thresholds increase continuously with increasing masker bandwidth over the entire range of bandwidths tested. Second, masking decreases with increasing signal frequency, which also agrees well with the data. The model slightly overestimates the amount of masking for the signal frequencies 16, 32, and 64 Hz for the narrow-bandwidth maskers ( $\leq 1$  octave) by about 1-3 dB.

Overall, the predictions demonstrate that second-order modulation bandpass filters account very well for the shape of the MTPs as well as for the masking data obtained in the band-widening experiment. The filter *shape* appears to play a larger role than the exact choice for the Q-value. A constant Q-value of 1 as an approximation to the best-fitting values still describes the data reasonably well.

## 2. Modeling modulation transfer functions

While the model parameters have been optimized to describe the amount of modulation masking correctly, it is not clear in advance whether the same model can also account for the shape of the TMTF, which serves as the reference condition for the masking data.

Figure 1.10 shows the TMTF for the 3-kHz wide noise carrier referred to as the broad-band carrier in the following. In addition to the data shown in Fig. 1.1, thresholds were also obtained at the intermediate signal frequencies 8, 32 and 128 Hz. The filled symbols indicate predictions obtained with the EPSM. Each simulated threshold represents the

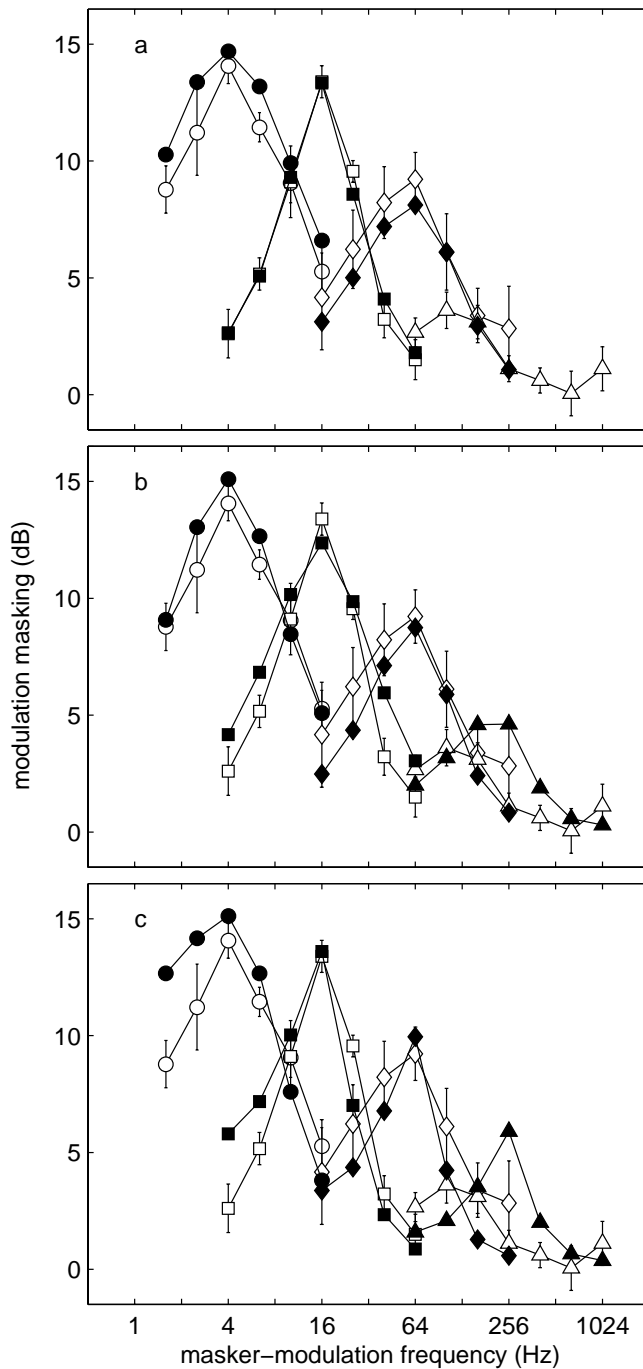


FIG. 1.7: Predicted masking patterns (filled symbols) in comparison with the experimental data (open symbols) from Fig. 1.1. Masking is plotted as the difference between the thresholds obtained in the masked and the reference condition. The experimental data are the same in each panel. Different realizations of the modulation filterbank were used for the simulations. The upper panel (a) represents model predictions obtained with the best-fitting  $Q$ -values for the modulation filters. The middle panel (b) shows simulations obtained with constant- $Q$ -bandpass filters ( $Q = 1$ ) as an approximation of the best-fitting filters from the upper panel. In the lower panel (c), the resonance filters proposed in earlier studies (Dau et al., 1997a,b, 1999) were used for the predictions.

envelope power of the carrier in the passband of the filter tuned to the signal frequency. As a consequence of the constant  $Q$ -value, the TMTF increases continuously at a rate of about 3 dB per octave with increases in signal frequency. This is in contrast to the empirical data which show a plateau up to about 16-32 Hz before threshold increases. The data are in line with those from many other studies (e.g., Rodenburg, 1972, 1977; Viemeister, 1977, 1979; Eddins, 1993; Strickland and Viemeister, 1997; Eddins, 1999)<sup>4</sup>.

<sup>4</sup>It should be noted that the current stimulus configuration represents a gated-carrier conditions. As was shown earlier (e.g., Viemeister, 1979; Sheft and Yost, 1990; Yost and Sheft, 1997), detection thresholds for very low modulation rates can be decreased when a continuous carrier is used instead of a gated one.

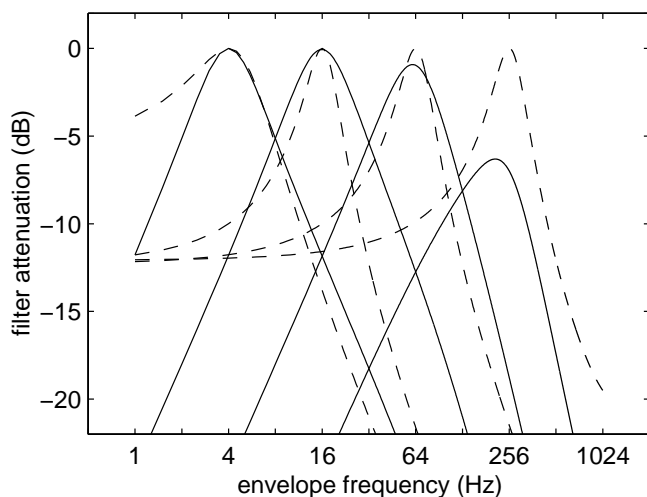


FIG. 1.8: Transfer functions of the second-order bandpass filters with constant  $Q = 1$  (solid functions) and of the first-order resonance filters suggested in earlier studies (dashed functions).

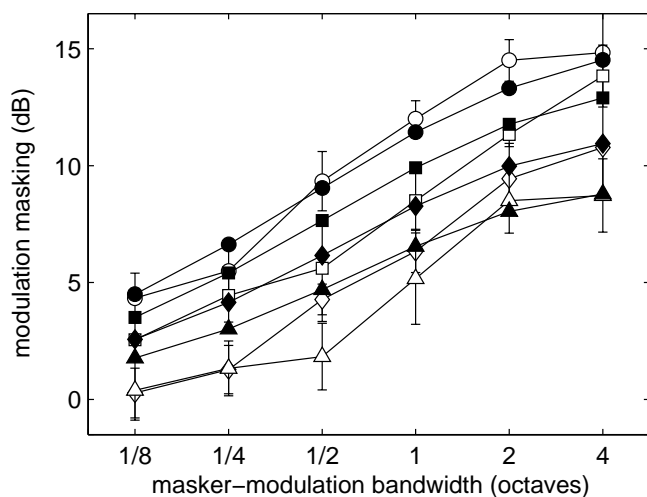


FIG. 1.9: Predicted masking (filled symbols) in the band-widening experiment. For direct comparison, the experimental data from Fig. 1.5 are replotted and indicated as open symbols.

To get a broader picture of the model's capabilities and limitations in predicting TMTFs, Fig. 1.11 shows data (open symbols) and model predictions (filled symbols) for a large set of carrier bandwidths ranging from 1 Hz to 6 kHz, with 6 kHz as the upper cut-off frequency of the carrier. The data are replotted from Dau *et al.*(1999). AM thresholds for 5 Hz (circles), 25 Hz (squares), and 100 Hz (diamonds) are plotted as a function of the carrier bandwidth. For a fixed carrier bandwidth, the figure indirectly reflects the principal shape of the TMTF (for details, see Dau *et al.*, 1999). That is, for large carrier bandwidths ( $\geq 250$  Hz) thresholds increase with increasing signal frequency; for small carrier bandwidths ( $\leq 25$  Hz) thresholds decrease with increasing signal frequency and for bandwidths in the region between 25 and 250 Hz threshold is largest for the intermediate

The subjects of the present study did not show the typical band-pass shape of the TMTF, with increased thresholds at very low modulation rates. Possibly, these subjects might show such an effect for rates lower than 4 Hz, the smallest rate tested in the present study. In any case, the basis for these differences still is controversial. According to Sheft and Yost (1990), threshold differences between the two presentation modes still occur for "fringe" durations of the carrier up to as much as 500 ms which can not be explained in terms of neural adaptation due to the carrier onset. Whatever the reason for these effects might be, we assume that the *amount of masking*, reflecting the difference between masked and reference threshold, roughly remains the same for both presentation modes.

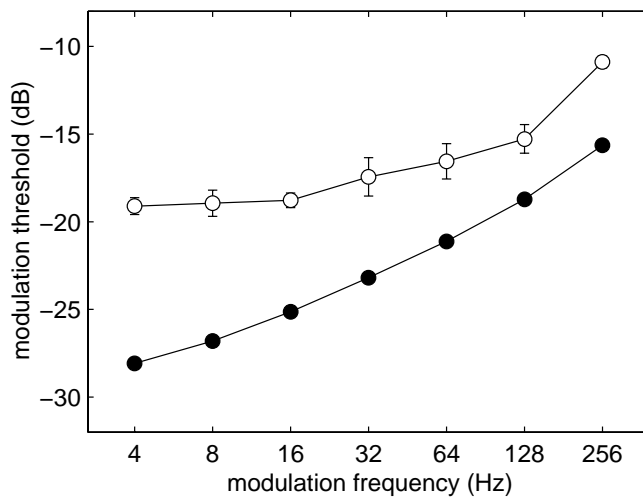


FIG. 1.10: TMTFs for broadband (1 to 4 kHz) noise as the carrier. The open symbols represent average experimental data, the filled symbols represent model predictions.

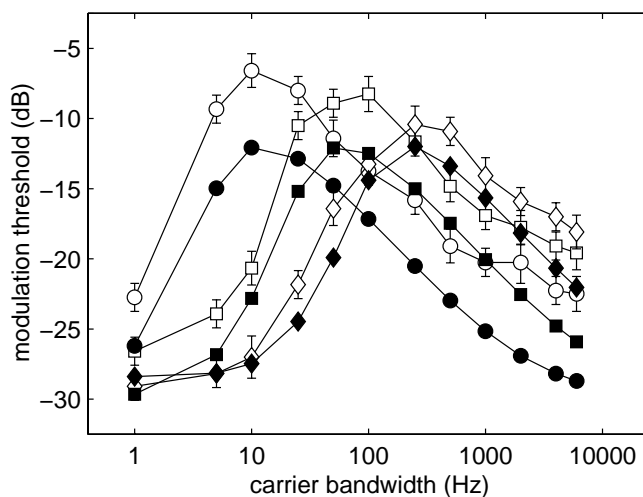


FIG. 1.11: AM thresholds for 5 Hz (circles), 25 Hz (squares), and 100 Hz (diamonds) as a function of the carrier bandwidth. The upper cutoff frequency of the carrier was 6 kHz. The open symbols indicate experimental data and are replotted from Dau *et al.* (1999). The filled symbols represent model predictions.

signal frequency (25 Hz).

The model accounts for the general shape of the threshold patterns. In particular, the shape of the predicted patterns is much closer to the data than that of the calculations presented in Dau *et al.* (1999), especially for small carrier bandwidths. This is a further indication of the appropriateness of the filter parameters chosen here. However, there remain some discrepancies between data and predictions. The predicted thresholds are generally lower than the corresponding experimental thresholds. The effect is small (1-3 dB) for a signal frequency of 100 Hz, but increases with decreasing signal frequency. For 5 Hz, the deviations between model and data amount to 3-8 dB. At large carrier bandwidths ( $\geq 250$  Hz) the model predicts a decrease of about 3 dB per octave increase of the carrier bandwidth, independent of signal frequency. In contrast, the experimental data show a smaller change in threshold (1-2 dB per octave) with increasing bandwidth beyond 250 Hz. Thus, the predicted TMTF for a wideband carrier (e.g., 3 kHz) is much too sensitive at 5 Hz with sensitivity decreasing too rapidly with increasing signal frequency. This was also observed in Fig. 1.10. The discrepancy between predicted and obtained TMTF will be discussed in detail in Sec. III C 1.

## B. Effects of peripheral filtering

### 1. Multi-channel model

The EPSM in its present state does not include the influence of peripheral filtering on modulation detection. Several studies suggested that AM thresholds do not depend much on the frequency region of a bandpass noise carrier. In other words, thresholds obtained with a fixed-bandwidth carrier centered at low frequencies are about the same as those for a carrier shifted towards higher frequencies (e.g., Viemeister, 1979; Eddins, 1993; Dau *et al.*, 1997b). These data would suggest that information about the signal modulation in the low-frequency region will be combined optimally across channels without loss of information. In fact, this would imply that the data could be modeled very efficiently by simply assuming one broad “predetection” filter as was already done by Viemeister (1979) as well as in the present model. However, even if such a simple description has some appeal, a more realistic model would be of more general interest. The “predetection” filter approach would not account for data in more complex conditions such as, e.g., modulation detection interference (MDI) (e.g., Yost and Sheft, 1988; Yost and Sheft, 1989; Yost *et al.*, 1989) or comodulation masking release (CMR) (e.g., Hall *et al.*, 1984; McFadden, 1986; Hall *et al.*, 1989) where across-peripheral-channel interactions seem to play a role. In the following this “predetection” filter EPSM is referred to as the single-channel EPSM.

In order to investigate the role of peripheral filtering on AM detection within the framework of the EPSM, a multi-channel version of the model was implemented, in which the envelope fluctuations are processed in all excited peripheral channels. Since the bandwidth of the envelope spectrum of a noise stimulus is directly related to the bandwidth of the stimulus’ audio-frequency spectrum, peripheral filtering produces narrowband envelope spectra. Assuming rectangular filter shapes for the peripheral filters, the corresponding modulation spectra can be derived analytically, as in the single-channel EPSM presented earlier. However, if one assumes more realistic filter shapes such as Gammatone filters (Patterson, 1987), the corresponding envelope spectra can no longer be easily derived analytically. In such a case, numerical simulations are needed where the envelope power spectrum is estimated for the output of each peripheral channel by averaging over a number of stimulus realizations.

In the following, sample TMTF simulations are presented, which were obtained with a multi-channel EPSM. The averaged envelope power of the carrier was calculated for each peripheral filter at the output of the modulation filter tuned to the signal frequency. Comparable to the single-channel model, AM threshold was estimated by adjusting the signal-modulation depth to match a 1-dB  $(S + N)/N$  decision criterion, corresponding to a signal-to-noise ratio of  $S/N = d'^2/2$  ( $\approx -6$  dB) for the combined observation. To satisfy this “overall” criterion within the multi-channel model, observations were combined across audio-frequency channels by assuming that

$$d' = \left( \sum_{i=1}^n (d'_i)^2 \right)^{1/2} = \left( \sum_{i=1}^n \frac{2S_i}{N_i} \right)^{1/2}, \quad (1.2)$$



where  $d'_i$  denotes the sensitivity index for peripheral channel  $i$ , which is proportional to the square root of the signal-to-noise ratio  $S_i/N_i$  in this channel (Tanner and Sorkin, 1972), and  $d'$  denotes the overall sensitivity index. This is equivalent to a linear combination of  $n$  independent observations according to the “integration model” by Green and Swets (1966). Thus, it is assumed within the multi-channel EPSM that information is combined without any loss, and that decisions are based on this combined information. Analogous to the analytical approach, resolution was finally limited by adding a -30 dB threshold.

As long as external noise is the limiting factor for detection, the independence of observations will depend on the correlation between the carrier waveform in the different peripheral channels. The correlation depends in turn on the shape and spacing of the filters. This was discussed in the framework of the study by Dau *et al.* (1997b). For instance, by allowing some overlap between adjacent filters, the observations in these filters are not statistically independent. By using a wider spacing, the information gain from combining a certain number of filters is larger. However, with a wider spacing, the number of filters that could be placed within a *given* spectral range would decrease. This decrease in the number of observations would affect thresholds more strongly than the gain from the statistical independence. Overall, it turned out that the typically chosen overlap at the -3 dB points of the transfer functions, which was also chosen here, allow a close to optimal detection of modulation imposed on a broadband noise carrier with a minimum computational load, i.e., for such a filter configuration, performance was comparable to that for nonoverlapping (rectangular) filters.

## 2. Predictions

Figure 1.12 shows predictions of the TMTF for broadband carriers. Four different model realizations were used. The dashed threshold curve represents results obtained with the original single-channel EPSM where thresholds were derived analytically. The filled circles represent single-channel predictions where thresholds were derived numerically from the averaged “real” envelope spectra instead of the approximated ones. The two curves differ by less than 1 dB. This small difference reflects the influence of bandwidth restriction of the carrier after modulation. While the numerical approach considers exactly the same stimulus conditions as in the real experiments, the analytical approach represents a (very good) approximation.

Results from the multi-channel predictions are indicated by the remaining symbols. The filled diamonds represent thresholds obtained with a model which assumes rectangular peripheral filters (within 1-4 kHz) with bandwidths according to the equivalent rectangular bandwidth (ERB) suggested by Moore and Glasberg (1983). For modulation frequencies below 32 Hz, peripheral filtering has no influence on thresholds. For frequencies larger than 32 Hz peripheral filtering starts having an influence. Thresholds are increased by 1, 3, and 8 dB for the frequencies 64, 128, and 256 Hz, respectively, relative to the single-channel simulation (filled circles). The reason for the large effect at high frequencies is that only a few peripheral channels (the upper 5 out of 11 filters in case of a modulation

frequency of 256 Hz) pass the signal modulation while the remaining channels in the lower carrier-frequency region infinitely attenuate the signal.

Results change considerably when Gammatone filters are used instead of the unrealistic rectangular filter shapes. Corresponding modeling results are indicated by the filled squares. Thresholds are decreased by about 1.5 dB relative to the single-channel simulation. This shift is caused by the partly correlated observations at the output of the overlapping Gammatone filters which are combined under the assumption of statistical independence within the model. Apart from this effect, peripheral filtering only has an effect at the highest signal frequency (256 Hz) where threshold is increased by about 2 dB.

Finally, the filled triangles in Fig. 1.12 represent predictions where, in addition, an instantaneous compressive nonlinearity was assumed for each Gammatone channel. A power-law function with an exponent of 0.4 was assumed. This is reasonably consistent with physiological estimates of the amount of compression on the basilar membrane for mid-range levels (Ruggero *et al.*, 1997). It has been shown that such a compressive nonlinearity can have major influence on masked thresholds as, e.g., in conditions of forward-masking (Oxenham and Moore, 1995; Moore and Oxenham, 1998; Oxenham and Plack, 1997; Plack and Oxenham, 1998). It also has been suggested that a peripheral nonlinearity affects the threshold for detecting a signal modulation in the presence of a pair of masker modulators beating at the rate of the signal frequency: the nonlinearity introduces a distortion component at the beat rate in the internal representation of the envelope which might interact with the signal component in a phase-sensitive way (Moore *et al.*, 1999). In each peripheral channel, the compressed stimulus  $c(t)$  was calculated from the input stimulus  $s(t)$  according to:

$$c(t) = \text{sign}(s(t)) \cdot |s(t)|^{0.4}, \quad (1.3)$$

where  $\text{sign}()$  denotes the signum function.

The negligible differences between thresholds obtained with (filled triangles) and without compression (filled squares) indicate that peripheral compression has no effect on the results in the framework of the current model. Since both signal modulation and inherent carrier modulations are compressed in the same way, the signal-to-noise ratio at the output of the model does not change. These predictions are consistent with the observation that sensorineural hearing impaired listeners often show about the same sensitivity for modulation independent of the amount of hearing loss (e.g., Bacon and Viemeister, 1985; Formby, 1987; Bacon and Gleitman, 1992; Derleth *et al.*, 2000), at least for narrow-band noise carriers, and also for broadband noise carriers as long as the hearing loss is relatively flat. Sensorineural hearing impairment is commonly associated with a loss or reduction of basilar-membrane compression (e.g., Yates, 1990; Glasberg and Moore, 1992; Ruggero and Rich, 1991).

## C. Discussion

The envelope power spectrum model (EPSM) accounts well for both modulation mask-

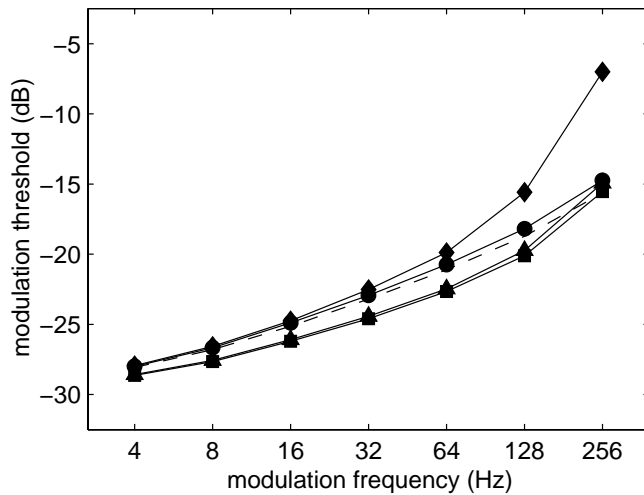


FIG. 1.12: Predicted TMTFs obtained with different versions of the EPSM. Dashed curve: Analytically derived thresholds obtained with the “single-channel” model as in the predictions from the previous figures. Filled circles: Numerically derived single-channel predictions based on averaged stimulus realizations (for details, see text). Filled diamonds: Numerically derived predictions obtained with the multi-channel EPSM assuming rectangular-shaped peripheral filters with bandwidths corresponding to the ERB values suggested by Moore and Glasberg (1983). Filled triangles: Corresponding multi-channel predictions assuming overlapping Gammatone filters. Filled squares: As for the triangles but, in addition, an instantaneous waveform compression was assumed in the preprocessing (for details, see text).

ing and modulation detection data. While the model parameters have been fitted to the masked threshold patterns (MTPs) the model also accounts for the data from the bandwidthing experiment as well as for the shape of the threshold patterns (from Fig. 1.11) for most of the carrier-bandwidth conditions. The use of second-order bandpass filters led to better results than the filters proposed in the earlier studies. The modulation filters appear to have a constant  $Q$ -value of about 1 for frequencies up to about at least 64 Hz. While frequency selectivity can clearly be observed up to this frequency, the effects are much weaker at higher frequencies. Within the model, an additional first-order low-pass filter with a cutoff frequency of 150 Hz introduces a certain sluggishness into the system which reduces the available dynamic range of detectable modulation depths at high frequencies. This corresponds, in principle, to the modeling strategy suggested by Viemeister (1979). However, Viemeister (1979) and Strickland and Viemeister (1997) suggested a much lower cut-off frequency (of 60-70 Hz) to fit their model to the TMTF data obtained with broadband carriers when a change in the envelope power was used as the decision statistic. A cut-off frequency of 150 Hz has also been suggested in a recent study on TMTFs and beat-detection thresholds for sinusoidal carriers (Kohlrausch et al., 2000).

### 1. The role of the decision variable

As described in Section 2., some discrepancies between predictions and data were observed for the shape of the TMTF obtained with a broadband noise carrier (Fig. 1.10). Due to the logarithmic scaling of the modulation filter bandwidth the predicted TMTF increases at a rate of about 3 dB per doubling of the signal frequency. Obviously, this

does not agree with the experimental data at low signal frequencies where threshold is constant up to about 30-60 Hz.

The question is what mechanisms may be involved which are not reflected in the EPSM. In the present state of the EPSM, modulation detection thresholds are directly related to the overall envelope power. It was assumed in the model that a 1-dB increase in power is required for detection. In terms of signal detection theory, this is equivalent to assuming that the envelope power (at the output of the modulation filter) forms a random variable, whose standard deviation and mean are proportional to one another. Such a decision statistic was also suggested by Viemeister (1979) who used the ac-coupled rms power of the output of a 64-Hz low-pass filter (instead of a bandpass modulation filter tuned to the signal frequency), calculated over the duration of the observation interval. Instead of the rms power, some studies have used the ratio of the largest to the smallest instantaneous value, max/min, (Forrest and Green, 1987), the fourth moment (Hartmann and Pumplin, 1988), the crest factor (Hartmann and Pumplin, 1988), or the average magnitude of the slope of the envelope (Richards, 1992) as the decision statistic (for details, see Strickland and Viemeister, 1996). Some models of modulation detection for noise carriers (e.g., Viemeister, 1979; Wakefield and Viemeister, 1990; Dau et al., 1997a,b) incorporate the “true” variation in the stimulus as the limiting quantity for detection, by using the same adaptive tracking procedure as in the experiments.

The use of a constant  $S/N$  or  $(S + N)/N$  ratio as the decision criterion, as assumed within the EPSM, appears to be in contrast to experimental data on modulation depth discrimination (Fleischer, 1980; Wakefield and Viemeister, 1990). In Fig. 3 of Wakefield and Viemeister (1990), discrimination thresholds are expressed as Weber fraction which directly corresponds to the  $S/N$  ratio defined in our study. They found small Weber fractions in the range from -4 to -7 dB for standards with modulation depths larger than -10 dB. A value of -6 dB equals a  $(S + N)/N$  ratio of 1 dB, as assumed in the present study. However, for modulation depths near detection thresholds, Weber fractions increased strongly with decreasing standard modulation depth corresponding to  $S/N$  ratios up to as much as 6 dB. Thus, Weber’s law does not hold over a large range of modulation depth. Wakefield and Viemeister (1990) were able to predict part of their discrimination data with Viemeister’s (1979) leaky integrator model (using a cascade of a bandpass filter with a 2-kHz bandwidth, a half-wave rectifier, and a time constant of 2.5 ms corresponding to a cut-off frequency of 64 Hz). Thresholds were estimated using a 2IFC tracking procedure which means that the statistical properties of the stimuli were taken into account. For low standard modulation depths, this model could predict the data reasonably well while it was much too sensitive at higher standard depths. In contrast, our  $S/N$  criterion would account for the data for high standard modulation depths, where Weber’s law holds, but would be much too sensitive at low standard depths. These findings suggest that a combination of both criteria might be appropriate for describing modulation detection and discrimination data.

In the following, model predictions are presented where decisions were based on the

*statistics* of the integrated envelope power. Each single component in the envelope spectrum of a bandpass Gaussian noise is known to closely follow an exponential distribution (Green *et al.*, 1992). This distribution has the specific property that the mean and standard deviation are equal. The mean of the envelope power passing through a specific modulation filter equals the sum over the means of a number of spectral components which are weighted by the (power) transfer function of the modulation filter. For a sufficiently large number of components, this integrated envelope power becomes a Gaussian random variable. The variance of this random variable can be computed by adding up the variances of the individual spectral components which were assumed to equal their squared means to a good first approximation<sup>5</sup>. It is assumed in the following that detection requires an increase in power which is proportional to three times its standard deviation. This value was chosen to closely reproduce the empirically found sensitivity to modulation. In addition, as an internal limit of resolution, it is assumed that the required increase in power can not become smaller than 1 dB, which corresponds to our original decision criterion. Again, an absolute threshold corresponding to a modulation depth of -30 dB was assumed. Figure 1.13 shows model predictions for three of the experimental conditions from the present study. The upper panel (a) of the figure shows results for the TMTF obtained with a broadband (1 to 4 kHz) carrier (compare to Fig. 1.10). The middle panel (b) shows results for the detection data from Fig. 1.11. Finally, the lower panel (c) represents simulations of the masked threshold patterns from Fig. 1.1 (compare to panel c in Fig. 1.7). The model accounts for several aspects in the data which were not accounted for by the original model. First, the slope of the broadband TMTF (upper panel) shows a 1.5-dB increase per octave increase of signal frequency for frequencies up to about 64 Hz, and an increase of 3-4 dB for larger frequencies. Thus the shape of the curve is more similar to that observed in the data. Predicted thresholds are still systematically lower than in the data, but the maximum deviation amounts to about 6 dB instead of 9 dB in case of the original model. Second, the model accurately accounts for the threshold patterns for 5, 25, and 100 Hz (except for very large bandwidths; see below). The statistical criterion leads to a decreasing peak modulation threshold with increasing modulation frequency, in accordance with the experimental data. Finally, the statistical approach also accounts for the shape of the MTP (lower panel) even though the modulation filters used in the simulations have been determined in combination with the original decision criterion. However, the amount of masking in the patterns can not be described correctly anymore. Masking is overestimated by as much as 8, 4, 3 and 1 dB for the signal frequencies 4, 16, 64, and 256 Hz, respectively.

Thus, it appears that the use of the statistical criterion allows an accurate description of the TMTF in various bandwidth conditions. In particular, the model predicts more realistic thresholds for low signal frequencies than the original approach does. At the

---

<sup>5</sup>We computed the envelope spectrum of a Gaussian noise carrier which was modulated by a bandlimited Gaussian noise (masker). The mean-to-sigma ratio of the envelope spectrum components, based on 10000 independent signal realizations, was close to unity.

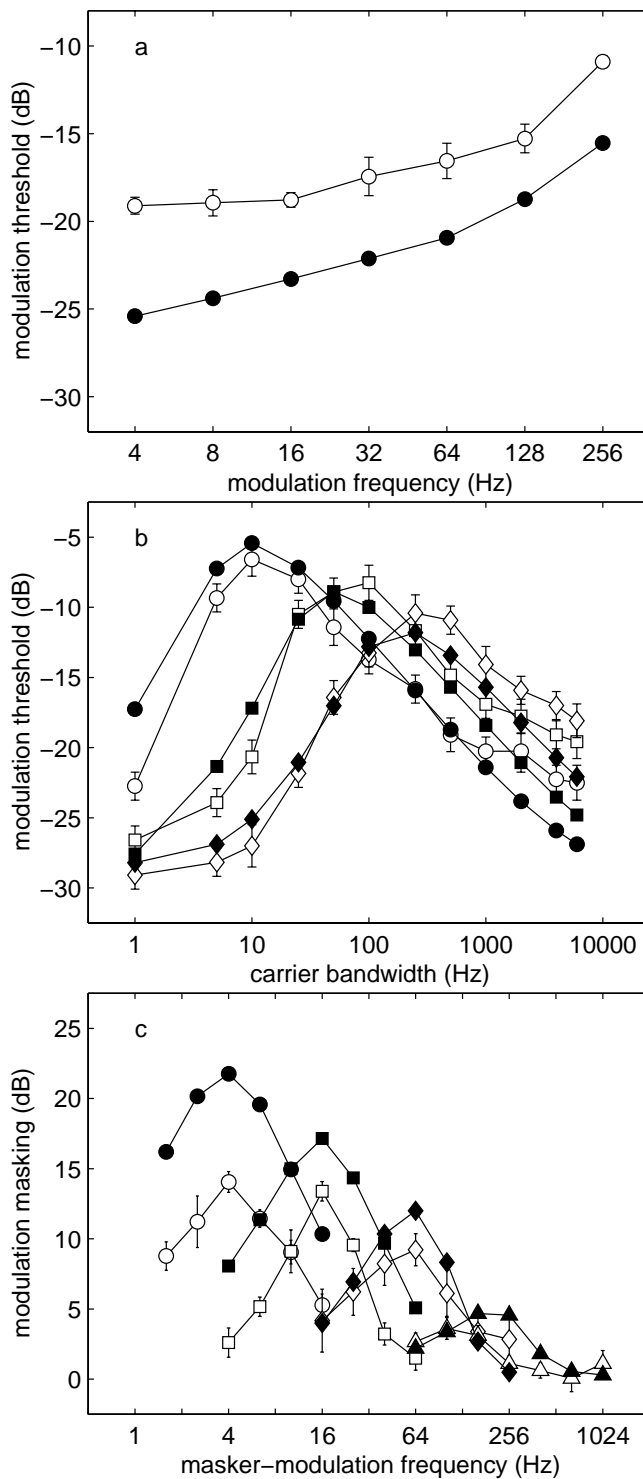


FIG. 1.13: Predictions (filled symbols) obtained with a “statistical” decision criterion. Decisions are based on the change of the standard deviation of the envelope power rather than on the change of the envelope power itself (for details, see text). Upper panel: Prediction of the TMTF for a broadband carrier (compare to Fig. 1.10). Middle panel: Prediction of the AM-detection data from Fig. 1.11. Lower panel: Prediction of the masking patterns derived from Fig. 1.1 (compare to Fig. 1.7). The open symbols indicate experimental data and are replotted from previous figures.

same time, however, the statistical approach predicts too much masking for these low frequencies when an additional modulation masker is presented. Based on these results it cannot be finally decided which criterion might be more appropriate. It should be noted, however, that the *shape* of the masking patterns can be described well – independent of the decision criterion.

## 2. Single-channel versus multi-channel model

Predictions obtained with the multi-channel model of the present study hardly differed from those obtained with the single-channel model. Within the multi-channel model peripheral filtering only has an effect on AM threshold when the signal is strongly attenuated by the transfer characteristic of one or more channel(s) assuming Gammatone filters as peripheral filtering stage. The model predicts an increase of threshold for 256 Hz of about 2 dB relative to the single-channel prediction while thresholds remain unchanged for lower signal frequencies. This was obtained for a noise carrier in the range from 1 to 4 kHz; however, the results may change somewhat when the frequency region of the carrier is shifted, since the transfer characteristic of the peripheral filters changes. Instantaneous waveform compression was shown to have no influence on the results since both signal and inherent modulation of the carrier are transformed in the same way resulting in the same signal-to-noise ratio at the output of the model. However, peripheral compression may have an effect in conditions of modulation masking produced by beating modulators. Moore *et al.* (1999) showed that the threshold for detecting 5-Hz signal modulation was affected by the presence of a pair of masker modulators beating at a 5-Hz rate. Their results suggested that such a nonlinearity introduces a distortion component in the internal representation of the envelope of the stimuli. The results could be fitted well using a model based on the concept of a modulation filterbank, following peripheral filtering, compression, and envelope extraction (Moore *et al.*, 1999). Compression might also influence results if one would assume a dynamic compression stage that acts differently for different modulation rates, as realized in the adaptation stage within the auditory processing model by Dau *et al.* (1996a,b, 1997a,b). Such a stage compresses low-frequency envelope components more than high-frequency components. Its transfer characteristic, however, is more complex than that of a linear high-pass filter (for details, see Dau *et al.*, 1999).

## 3. Combination across peripheral filters

The present model assumes optimal combination of signal information across frequency. Independent of whether decisions were based on the envelope power or on its standard deviation, the model predicts a 3-dB decrease of threshold per doubling of the carrier bandwidth, as observed for large bandwidths ( $\geq 250$  Hz) in Fig. 1.11 and Fig. 1.13 (b), respectively. This differs from the data which only show a smaller decrease of 1-2 dB. The conditions at the higher carrier bandwidths reflect situations in which several peripheral channels are excited by the stimulus while for smaller bandwidths activity is mainly restricted to the peripheral channel at or near 6 kHz. It appears that the real observer can combine signal information from different auditory filters but only less effectively than an ideal observer. Except for the study by Strickland and Viemeister (1997) who found a 3-4 dB effect, our data are consistent with several studies where smaller effects were observed (Maiwald, 1967; Viemeister, 1979; Eddins, 1993, 1999; Strickland, 2000). For example, Eddins (1999) and Strickland (2000) showed that the TMTF for bandpass noise

centered at low frequencies is shifted towards higher thresholds relative to carriers (of the same bandwidth) centered at high frequencies. Furthermore, Eddins (1999) observed an increase in sensitivity for 8 Hz SAM of about 6 dB as the carrier bandwidth increased from 400 to 3200 Hz, corresponding to a 2 dB change per doubling the carrier bandwidth. Optimal detection would lead to a 9 dB difference. Thus, the results obtained in the present study as well as those from the other studies might be accounted for by assuming a sub-optimal integration process across different peripheral channels. This observation is consistent with results from van den Brink and Houtgast (1990a,b) who showed that across-frequency integration in signal detection is *less* efficient for long-duration ( $> 100$ ms) broadband stimuli than for short-duration ( $< 100$  ms) broadband signals presented in a broadband noise masker. It appears that similar mechanisms play a role in modulation detection conditions. In the present study, the signal modulation (which is coherent across frequency) had a duration of 500 ms and deviations from optimal performance were observed for broadband carriers.

## V. SUMMARY AND CONCLUSIONS

- Peaked masked-threshold patterns (MTPs) with about the same shape on a logarithmic frequency scale were obtained for signal modulation frequencies from 4 to 64 Hz using half-octave-wide noise as the modulation masker imposed on the same noise carrier as the signal. The shape of the MTPs did not change markedly when the carrier band was shifted towards higher frequencies. The shape of the MTPs corresponded well with the shape of psychophysical tuning curves (PTCs) obtained with the same masker modulation and carrier. The data obtained in the band-widening experiment also suggested a constant relative scaling of the modulation filter bandwidth but did not allow any assumptions about the underlying filter shape.
- Within the framework of the envelope power spectrum model (EPSM) a modulation filterbank consisting of second-order bandpass filters and an additional first-order lowpass filter at 150 Hz was fitted to the data. The AM threshold was estimated by calculating the envelope power of the carrier and masker modulation passing the modulation filter tuned to the actual signal frequency. A constant Q-value of 1 for the bandpass filters accounted very well for the masking data. The overall lowpass filter accounted for the general decrease in sensitivity for frequencies larger than about 150 Hz.
- The same model also accounted reasonably well for the shape of the TMTF obtained for carrier bandwidths in the range from 1 to 6000 Hz. However, for low signal frequencies the model generally underestimated threshold in all bandwidth conditions by 3 to 5 dB. In addition, for large carrier bandwidths the model generally underestimated threshold for all signal frequencies tested. Two possible explanations for these deviations were discussed. First, a decision criterion based on the standard



deviation of the envelope power (instead of the envelope power itself) was shown to account very well for the TMTF for low signal frequencies. Second, a “sub-optimal” signal-integration process across carrier frequency will probably account for the data for large carrier bandwidths.

- Peripheral filtering influenced modeling results only in conditions when the signal modulation was strongly attenuated by the transfer function of the filters. Using Gammatone filters within the multi-channel EPSM, for a carrier ranging from 1 to 4 kHz, threshold was unaffected for signal frequencies up to 128 Hz and was increased by 2 dB for 256 Hz, relative to the predictions without peripheral filtering. Instantaneous waveform compression at a peripheral level did not have any effect on the simulations in the conditions tested in the present study.

## **ACKNOWLEDGMENTS**

We thank Sid Bacon and two anonymous reviewers for very helpful comments on an earlier version of the manuscript. We also want to thank Birger Kollmeier, Jesko Verhey, Brian Moore, and Andrew Oxenham for critical remarks and suggestions. This work was supported by the Deutsche Forschungsgemeinschaft and the Max Kade Foundation.



## CHAPTER 2

# EXTERNAL AND INTERNAL LIMITATIONS IN AMPLITUDE-MODULATION PROCESSING<sup>a</sup>

### ABSTRACT

Three experiments are presented to explore the relative role of “external” signal variability and “internal” resolution limitations of the auditory system in the detection and discrimination of amplitude modulations (AM). In the first experiment, AM-depth discrimination performance was determined using sinusoidally modulated broadband-noise and pure-tone carriers. The AM depth,  $m$ , of the standard ranged from  $-28$  to  $-3$  dB (expressed as  $20\log_{10}(m)$ ). It was found that AM-depth discrimination thresholds are a fraction of the AM depth of the standard for standards down to  $-18$  dB, in case of the pure-tone carrier, and down to  $-8$  dB, in case of the broadband-noise carrier. For smaller standards, AM-depth discrimination thresholds were constant on an absolute scale. In the second experiment, AM-detection thresholds were obtained for signal-modulation frequencies of 4, 16, 64, and 256 Hz, applied to either a band-limited, random-noise carrier or a deterministic (“frozen”) noise carrier, as a function of carrier bandwidth (4 to 2048 Hz). In general, detection thresholds were higher for the random- than for the frozen-noise carriers. For both carrier types, the threshold followed the pattern expected from frequency-selective processing of the stimulus envelope. The third experiment investigated AM masking at 4, 16 and 64 Hz in the presence of a narrowband masker modulation. The variability of the masker was changed from entirely frozen to entirely random, while the long-term average envelope power spectrum was held constant. The experiment examined

---

<sup>a</sup>This chapter is submitted as S. D. Ewert and T. Dau, (2002). “External and internal limitations in amplitude-modulation processing” to J. Acoust. Soc. Am. .

the validity of a long-term average quantity as the observation variable, and the role of memory in experiments with frozen-noise maskers. The empirical results were compared to predictions obtained for two modulation-filterbank models. The predictions revealed that AM-depth discrimination and AM detection are limited by a combination of the external signal variability and an internal “Weber-fraction” noise process.

## I. INTRODUCTION

Intensity discrimination represents one of the basic measures in psychoacoustics. Discrimination functions for the stimulus intensity provide information about the fidelity of intensity coding in the auditory system. Generally, two separate pulses of sound are presented successively, one being more intense than the other, and the subject is required to indicate which was the more intense. For pure-tone (deterministic) stimuli, the smallest detectable change in intensity  $\Delta I$  is, to a good first approximation, a constant fraction of the intensity  $I$  of the stimulus (Riesz, 1928; Florentine, 1983; Florentine *et al.*, 1987). Thus Weber’s law, stating that  $\Delta I/I$  is constant, roughly describes listeners performance; the quantity  $\Delta I/I$  is referred to as the Weber fraction. Empirically obtained Weber fractions amount to about 1 dB. For deterministic stimuli, a constant Weber fraction can generally be modeled by the assumption that there is “internal noise”, or variability, in the internal (observation) variable reflecting stimulus intensity, proportional to the mean of the variable. This variability is intended to reflect an inherent property of the internal coding mechanism. In terms of signal detection theory (Green and Swets, 1966) this assumption is necessary to account for the discrimination performance of the real observer (subject) in comparison to that of the ideal observer. For deterministic stimuli, such as pure tones, the detection performance of the ideal observer would be determined by the amount of change, independent of the stimulus intensity, since the stimuli exhibit no “external” variability. If stochastic stimuli are used, such as random noise, “external noise” comes into play, in addition to the internal variability of the coding process. If the amount of external noise, introduced by the stimulus, exceeds the amount of internal noise, this will affect the observer’s detection performance. In case of intensity discrimination, broadly similar results have been obtained for deterministic pure-tone and broadband “pseudo-random” as well as for random stimuli (Florentine *et al.*, 1987, Raab and Goldberg, 1975), indicating that the internal noise represents the main determinant. When intensity discrimination for a pure tone is measured in the presence of a masker noise, the effect of the masker is negligible when the pedestal is more than 5-10 dB above its threshold (Jesteadt *et al.*, 1977). However, slight changes in performance are observed for pedestals up to 15-20 dB above masked threshold due to the external variability of the masker (Henning and Psotka, 1969).

The present study is concerned with the coding mechanisms and the role of internal and external noise in amplitude-modulation (AM) processing. Amplitude modulations are a common feature of sounds in our natural surroundings. The temporal structure of the

envelope of a stimulus carries important information, such as in human speech, animal sounds, and environmental sounds. Several studies have investigated the limitations of the auditory system to resolve AM. For example, Fleischer (1980) obtained AM-depth discrimination functions for sinusoidal AM of pure-tone carriers, for standard modulation depths in the range from -14 to 0 dB (in units of  $20 \log m$ , where  $m$  is the modulation depth). For a similar limited range of standards (-12 to 0 dB), Ozimek and Sek (1988) measured discrimination thresholds for modulated octave-band noises. Modulation frequencies in the range from 1 to 64 Hz were tested, imposed on carriers with center frequencies in the range from 250 to 2000 Hz. Wakefield and Viemeister (1990) studied AM-depth discrimination for modulation rates of 25, 100, and 400 Hz, using broadband noise carriers. In the comparable range of standard modulation depths, their results were in agreement with the findings of Fleischer (1980) and Ozimek and Sek (1988). More recently, Lee and Bacon (1997) investigated AM-depth discrimination for a sinusoidal carrier as a function of stimulus duration. For the longest signal durations, 400 and 800 ms, their results were also broadly compatible with the findings from the other studies mentioned above: A roughly constant relative AM-depth discrimination threshold, corresponding to a roughly constant Weber fraction, was found in all conditions where the standard AM depth was greater than about -15 dB, i.e., well above detection threshold. On average, the AM-depth discrimination threshold was slightly greater than 1 dB in these conditions, independent of carrier type.

Wakefield and Viemeister (1990) and Fassel (1995) also obtained discrimination thresholds for standard depths near the absolute AM-detection threshold. For small standard depths below -15 dB, Wakefield and Viemeister found a region of roughly constant absolute AM-depth discrimination thresholds, using a broadband-noise carrier. Similarly, Fassel (1995) found an essentially constant absolute discrimination threshold, using a sinusoidal carrier. However, he observed a constant absolute threshold throughout the entire range of standards tested (-40 to -3 dB), in contrast to the studies by Wakefield and Viemeister (1990) and Fleischer (1980). Whether the differences between the studies were due to differences of the carrier type (broadband noise versus pure tone), or due to other factors, has not been clarified yet.

In Wakefield and Viemeister (1990), the leaky-integrator model proposed by Viemeister (1979) was tested in conditions of AM-depth discrimination. The model was originally designed to account for the temporal modulation transfer function (TMTF) obtained with a broadband noise carrier. It consists of a 2-kHz-wide bandpass filter, a half-wave rectifier, and a single-pole low-pass filter with a time constant of 2.5 ms, representing a modulation-low-pass filter with a cutoff frequency of about 64 Hz. The detection device is based on the comparison of the rms (root-mean-square) envelope power at the output of the modulation-low-pass filter in a simulated alternative forced-choice procedure. When this model is applied to two noise stimuli, one modulated at a certain rate and standard modulation depth, the other one modulated at the same rate but higher modulation depth, the model is able to correctly discriminate between the two stimuli if the increase in rms

power at the output of the modulation low-pass, caused by the increase in AM depth, exceeds the random fluctuations of the rms power due to the stochastic nature of the noise carrier. The model was shown to accurately account for discrimination data when the AM depth of the standard was below  $-10$  dB. However, for greater standard depths of  $-7.5$  and  $-5$  dB, predicted thresholds were considerably lower (better performance) than the data. The reason for this discrepancy between data and model predictions might have been that the discrimination performance within the model is entirely determined by the external variability of the noise carrier. This results in a constant absolute discrimination threshold, independent of the standard AM depth of the noise. The model therefore fails for large standard depths where a constant Weber fraction is observed (corresponding to an increase of the absolute discrimination threshold with increasing standard depth). Moreover, the model would predict infinitely small discrimination thresholds in conditions with deterministic stimuli, such as sinusoidal or frozen-noise carriers, because it does not make any assumptions about the internal limitations inherent in the encoding of AM.

In the present study, three experiments are presented to explore the relative role of internal and external limitations in auditory AM processing. The first experiment examines AM-depth discrimination with a (random) broadband-noise carrier and a (deterministic) pure-tone carrier. Thus, this experiment investigates the influence of external noise (introduced by the random-noise carrier) and the role of internal noise (representing the only limitation in the case of the pure-tone carrier), using a single group of subjects and experimental setup. The second experiment compares AM-detection thresholds obtained for a random-noise and a frozen-noise carrier, as a function of the carrier bandwidth. This experiment addresses the question to what extent AM detection can be explained in terms of the average envelope power only, or in terms of a combination of the average envelope power and the variability of this value. In the third experiment, AM masking is investigated. The detection of a sinusoidal signal modulation is measured in the presence of a narrowband-noise masker modulation, centered at the signal-modulation frequency. The degree of variability of the noise masker is changed as a parameter such that conditions ranging from entirely frozen to entirely random are tested. The long-term average envelope power spectrum remains constant across conditions. This is done in order to examine the validity of a long-term average quantity as the observation variable, and to investigate the role of memory in experiments with frozen stimuli.

The empirical data of the present study are compared to predictions of the signal processing model by Dau *et al.* (1997a,b), in the following referred to as the “perception model” (PEMO), and the envelope power spectrum model according to Ewert and Dau (2000), in the following referred to as the EPSM. Constraints on future models of AM processing are discussed.

## II. MODELS OF AM PROCESSING

In the following, the two processing models used in the present study are briefly de-

scribed. For a detailed description, the reader is referred to the respective publications.

The first model used in the present study, the envelope power spectrum model, EPSM, has a structure similar to Viemeister's (1979) leaky-integrator model, but assumes a modulation filterbank instead of the low-pass filter. The model has proven successful in various conditions of AM detection and AM masking. It consists of only three stages: Hilbert-envelope extraction of the stimuli, a modulation filterbank, and a decision stage. The modulation-filterbank design was chosen according to Ewert and Dau (2000). Only the filter tuned to the signal modulation is considered. Effects of peripheral bandpass filtering and adaptation are not taken into account. Within EPSM, decisions are based on the mean integrated envelope power and the standard deviation of this value at the output of the modulation filter. It is assumed that the detection of the signal requires an increase in power that is proportional to its standard deviation. It is further assumed that the required increase in power cannot be less than 1 dB relative to the mean power, which reflects the "classical" constant Weber fraction. In addition, an absolute lower limit of resolution corresponding to a modulation depth of -30 dB is assumed.

The second, more complex perception model, PEMO, was designed to account for detection data in various experimental conditions. It has proven successful in predicting data from spectral and spectro-temporal masking (Verhey *et al.*, 1999; Derleth and Dau, 2000; Verhey, 2002), non-simultaneous masking (Dau *et al.*, 1996a,b), and modulation detection and modulation masking (Dau *et al.*, 1997a,b; Dau *et al.*, 1999). The model consists of the following basic stages: peripheral filtering, envelope extraction, non-linear adaptation, modulation filtering and an optimal detector as the decision device. To simulate the bandpass characteristic of the basilar membrane, the gammatone filterbank by Patterson *et al.* (1987) is used. At the output of each peripheral filter, the model includes half-wave rectification and low-pass filtering at 1 kHz, i.e., for high center frequencies, this stage essentially preserves the envelope of the signal. Effects of adaptation are simulated by a non-linear adaptation circuit (Püschel, 1988, Dau *et al.*, 1996a). For a stationary input stimulus, this stage realizes a compression close to logarithmic. With regard to the transformation of signal envelope fluctuations, the adaptation stage transforms rapid input fluctuations linearly. The stimuli at the output of the adaptation stage for each channel are then processed by a linear modulation filterbank. In the present study, only the output of the modulation filter tuned to the signal-modulation frequency is considered. When considering modulation processing, the most relevant difference between PEMO and EPSM is reflected in the decision criterion. Limitations of resolution are simulated by adding a constant-variance internal noise to the modulation-filter output. The variance of the internal noise was adjusted to account for a lower resolution limit of about -30 to -40 dB for AM-detection with frozen-noise carriers. The decision device is realized as an optimal detector. Decisions are based on the cross correlation between the internal representation of the current stimulus and a template, which is a stored representation of a stimulus with a supra-threshold signal.

When tested in modulation-detection conditions with noise carriers, the detection per-

formance of the model is mainly limited by the statistical properties of the stimulus. So far, the model has not been tested explicitly in conditions of AM-depth discrimination. Even though the model’s preprocessing stages and the decision device differ considerably from Viemeister’s (1979) leaky-integrator model, described in the introduction, it can be expected to predict the same constant absolute AM-depth discrimination threshold when applied to broadband noise carriers. In contrast to the Viemeister model, PEMO includes an internal-noise source of constant variance, that limits the resolution in conditions without external variability of the stimuli.

### III. AM-DEPTH DISCRIMINATION WITH NOISE AND PURE-TONE CARRIERS

#### A. Method

##### 1. Subjects

Five normal-hearing subjects participated in the experiment. Their age ranged from 23 to 36 years. Three subjects (MB, TD, and SE) had experience in other psychoacoustic experiments. Subjects SE and TD were the two authors. Two subjects (FM and CS) had no previous experience in psychoacoustic experiments and received at least three hours of training before data collection started. Subjects MB, FM and CS were paid for their participation on an hourly basis.

##### 2. Apparatus and stimuli

The subjects were seated in a double-walled sound attenuating booth and listened diotically via AKG K-501 headphones. Signal generation and presentation during the experiments were computer controlled using the AFC software package for MATLAB, developed at the University of Oldenburg. The stimuli were digitally generated at a sampling rate of 48 kHz and converted to analog signals by a two-channel 24-bit DAC including reconstruction filtering (SEKD ADSP 2496). The transfer function of the headphones was digitally equalized (64 point FIR filter) to match a flat amplitude response between 0.1 and 20 kHz, measured with the artificial ear (B&K 4153). A sinusoidal AM of 16 Hz was applied to a broadband-noise or a pure-tone carrier. The noise carrier had a bandwidth of 2048 Hz and an upper cutoff frequency of 4 kHz. The pure-tone carrier had a frequency of 4 kHz. Modulations were applied during the entire carrier duration of 500 ms. The stimuli were gated with 50-ms  $\cos^2$  onset and offset ramps.

The stimuli were defined as follows:

$$s(t) = A [1 + m_s \sqrt{1 + m_{inc}} \sin(2\pi f_m t)] c(t), \quad (2.1)$$

where  $m_s$  is the standard modulation depth,  $m_{inc}$  the relative AM increment and  $c(t)$  represents the carrier waveform. The signal-modulation frequency,  $f_m$ , was 16 Hz. In the



two reference intervals,  $m_{inc}$  was set to zero. The standard modulation depth was varied in 5 dB steps, ranging from  $-28$  to  $-3$  dB. To avoid possible level cues due to the presence of the signal modulation, the energy increase caused by the amplitude modulation was compensated for. The overall presentation level was 65 dB SPL.

### 3. Procedure

A three-interval, three-alternative forced-choice paradigm was used to measure AM-depth discrimination functions. A two-down one-up procedure was used to estimate the 70.7 % correct point of the psychometric function (Levitt, 1971). Subjects had to identify the one randomly chosen interval containing the increased AM depth of  $m_c = m_s \sqrt{1 + m_{inc}}$ . The standard AM depth in the two reference intervals was  $m_s$ . The fractional increment  $m_{inc}$ , corresponding to the Weber fraction  $(m_c^2 - m_s^2)/m_s^2$ , was varied in dB ( $10 \log m_{inc}$ ). The three observation intervals were separated by 500 ms silent intervals. The threshold run started at  $10 \log m_{inc} = 0$  dB. The initial step size was 2 dB and was divided by 2 after every second reversal until it reached the final stepsize of 0.5 dB. Six reversals were obtained at the final stepsize and the threshold estimate was taken as the mean value of  $10 \log m_{inc}$  at these reversals. Reported thresholds represent the mean of the estimates from four runs.

As already reported in Wakefield and Viemeister (1990), sequential dependencies among conditions can strongly influence AM-depth discrimination thresholds. Comparable with Wakefield and Viemeister's observations, our own preliminary experiments revealed that, particularly for low standard modulation depths, a much higher increment was required for detection when the task followed a run with a high standard modulation depth than when it followed a run with a lower or only slightly higher standard modulation depth. Following the findings of Wakefield and Viemeister (1990), AM-depth discrimination thresholds were measured in three blocks of two randomly mixed standard modulation depths separated by only 5 dB ( $-3$  and  $-8$ ,  $-13$  and  $-18$ ,  $-23$  and  $-28$  dB). For each subject, the three blocks were presented in random order.

## B. Results and discussion

The pattern of results was similar for the five subjects, so the mean data and standard deviations are shown. Figure 2.1 shows AM-depth discrimination thresholds, expressed as Weber fractions, as a function of the standard AM depth. The left panel shows results for the broadband-noise carrier, the right panel shows results for the pure-tone carrier. For the highest standard depth (right-most data point in each panel), the Weber fraction is similar for the two carrier types ( $-4$  dB for the noise, and  $-5.2$  dB for the tone carrier). For standard depths below  $-8$  dB, the Weber fraction rises continuously with decreasing standard depths in the case of the noise carrier. In contrast, in the case of the pure-tone carrier, the Weber fraction tends to stay constant for standard depths down to  $-18$  dB. For smaller standards, the Weber fraction increases at approximately the same rate as in

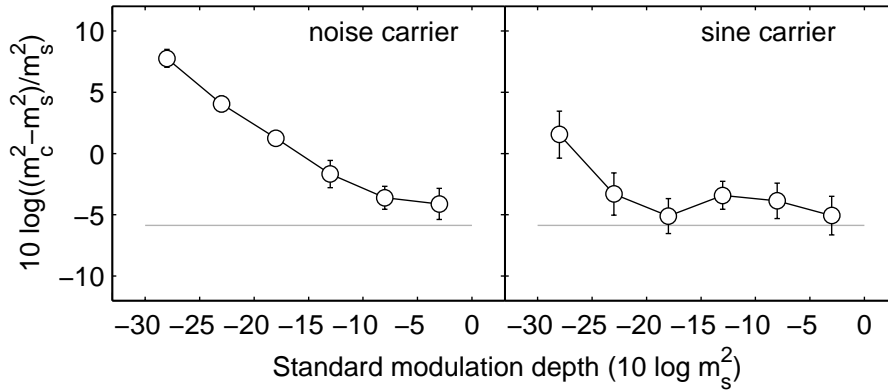


FIG. 2.1: Weber fractions for AM-depth discrimination as a function of the AM depth of the standard. The left panel shows data for a 2048-Hz-wide random-noise carrier with an upper cutoff frequency of 4 kHz. The right panel shows data for a 4-kHz sinusoidal carrier. The horizontal line indicates the expectation for a 1-dB increase required for discrimination.

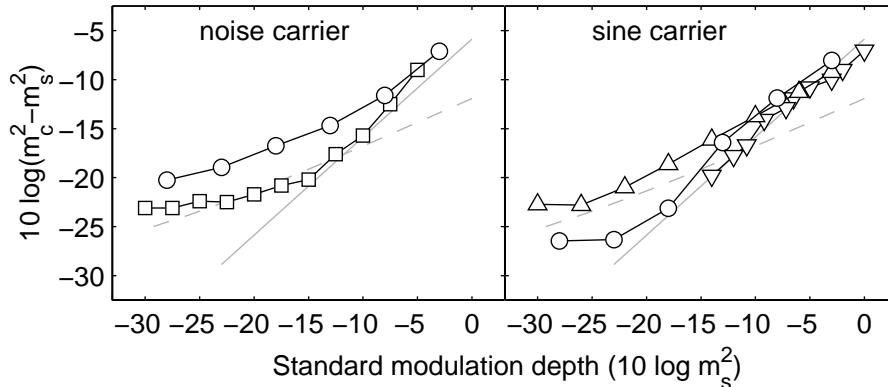


FIG. 2.2: Comparison of the data from Fig. 2.1 (circles), replotted as AM-depth discrimination functions, and data from the literature. The left panel (noise carrier) shows data by Wakefield and Viemeister (1990) as squares. In the right panel (sinusoidal carrier) the downward pointing triangles show data by Fleischer (1980) while the upward pointing triangles represent data by Fassel (1995). The solid line indicates the expectation for a constant 1-dB increase (Weber fraction) required for discrimination. The dashed line indicates the slope for a constant absolute increment.

the case of the noise carrier. For the lowest standard depth (-28 dB), the Weber fraction is 2 dB for the pure-tone carrier, compared to 7.5 dB for the noise carrier. A constant 1-dB increment in AM-depth discrimination (constant  $(S+N)/N$ ) would correspond to a value of -5.9 dB, if expressed as a Weber fraction. This value is indicated by the horizontal line in Fig. 2.1. At high standard depths, the obtained Weber fractions in the data would correspond to a just-detectable AM-depth increment of about 1.4 dB for the noise carrier and 1.2 dB for the pure-tone carrier.

In Fig. 2.2, the data from Fig. 2.1 are replotted as discrimination function  $10 \log(m_c^2 - m_s^2)$  and indicated by the circles. In addition, the data by Wakefield and Viemeister (1990) for a noise carrier (left panel, squares) and the data by Fassel (1995) and Fleischer (1980) for sinusoidal carriers (right panel, upward and downwards triangles) are shown. The

discrimination function that would correspond to a constant 1-dB increment is indicated by the solid line. The expectation assuming a constant absolute increment is represented by the dashed line. Altering the magnitude of either the relative or the absolute increment would result in a (vertical) displacement of the corresponding lines, while maintaining the slope. At high standard depths, the measured threshold function from the present study approaches the 1-dB-increment criterion (solid line) while, at small standard depths, it has about the same slope as the expected curve assuming the constant absolute-increment criterion (dashed line). The transition occurs at standard depths of about -8 dB for the noise carrier and at about -18 dB for the pure-tone carrier. For the noise carrier, the existence of the two different characteristic regions seems even more distinct in the Wakefield and Viemeister (1990) data. Compared to the current data, they found generally lower (better) discrimination thresholds, except for the highest standard depth of -5 dB. The reason for the differences between both sets of data is not clear, since the experimental paradigm used in the two studies is the same. One explanation might be the larger carrier bandwidth used in their study: The noise carrier had a bandwidth of 10 kHz, compared to a bandwidth of only 2048 Hz in the present study. Another difference is that in the present study, the Weber fraction  $m_{inc}$  was directly adjusted during a threshold run (and was dependent on the standard depth) while Wakefield and Viemeister (1990) used a constant stepsize, independent of the standard modulation depth, resulting in much larger steps for small standards than in the present study. However, in spite of these differences, both studies clearly show the two different slope regions in the AM-depth discrimination data with broadband-noise carriers.

For the pure-tone carrier (right panel of Fig. 2.2), the data of the present study are similar to the data of Fleischer (downward triangles): For standards depths above -18 dB, both data sets closely match the 1-dB-increment expectation (solid line). This is also the case for the data by Fassel (upward triangles). However, for standard depths below -15 dB, he obtained considerably higher (worse) discrimination thresholds. In this case, the threshold function by Fassel exhibits about the same slope as the expected function for the constant absolute increment (dashed line). The discrepancy between the data by Fassel (1995) and the other data might be explained by the fact that he applied a Hanning window over the entire duration of the modulation, whereas, for example, in the present study only the first and last 50 ms of the 500-ms modulation were windowed. At small AM depths, this probably results in a reduced effective observation interval for the modulated part of the stimulus, since only the central part of the Hanning window contributes to detection. This may have caused the increased discrimination thresholds for the smallest standard depths (Lee and Bacon, 1997).

Figure 2.3 shows model predictions (closed symbols), obtained with the two AM-processing models described in Section II., in comparison to the experimental data (open symbols) replotted from Fig. 2.1. The closed circles represent predictions obtained with the envelope power spectrum model (EPSM), while the closed squares show results for the more complex perception model (PEMO). For the noise carrier, shown in the left

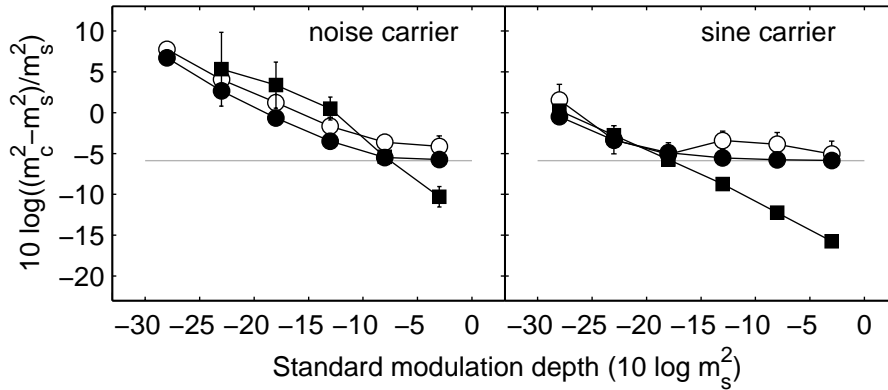


FIG. 2.3: Model predictions (closed symbols) and empirical data (open symbols, replotted from Fig. 2.1) for AM-depth discrimination. The closed circles represent predictions obtained for the EPSM, the closed squares represent PEMO predictions. For the lowest standard AM depth (-28 dB), no prediction could be obtained for PEMO within the upper limit (9 dB) of the tracking procedure.

panel, the EPSM predictions follow the empirical data very nicely. The prediction shows a constant Weber fraction for standard depths greater than  $-8$  dB while below  $-8$  dB the model predicts an increase of the Weber fraction with decreasing standard depth. The maximum deviation between data and prediction is less than 3 dB. The other model, PEMO, indicated by the closed squares, predicts a continuously changing Weber fraction throughout the entire dynamic range of standard depths. For standard depths below  $-10$  dB, the slope of the predicted curve roughly follows the slope in the data. In this region, the predicted values for the Weber fraction are slightly shifted towards higher values compared to the data and the EPSM predictions. For standard depths above  $-8$  dB, the model clearly overestimates the experimental performance (i.e., underestimates the measured Weber fractions). The differences between the two models are larger in the right panel of Fig. 2.3 where the data for the pure-tone carrier are shown. Here, for standard depths below  $-18$  dB, both models show excellent agreement with the data while, for higher standard depths, the predictions strongly differ from each other. In this region, EPSM correctly accounts for the constant Weber fraction observed in the data, while PEMO strongly overestimates the listeners performance.

In the framework of the models, the two characteristic regions in the data can be explained in terms of two different resolution limitations. One limitation is related to the external variability of the stimuli. A second limitation is related to the properties of the internal coding process. PEMO is limited by the external fluctuations inherent to the noise carrier or, in case of the sinusoidal carrier, by the assumption of a constant variance internal-noise process. Both limitations are independent of the standard modulation depth. These two limitations can explain the region of decreasing Weber fractions as observed for small standard modulation depths in the data. They overestimate, however, the discrimination performance for large modulation depths of the standard as it is also the case for the leaky-integrator model tested in Wakefield and Viemeister (1990). EPSM adds the assumption of an internal-noise process with a variance equal to the value of the

observation variable. With this assumption, EPSM is able to account for the region of constant Weber fractions observed in the data.

## IV. AM DETECTION WITH FROZEN- AND RANDOM-NOISE CARRIERS

### A. Rationale

In the previous experiment, AM-depth discrimination was investigated using a random-noise carrier and a deterministic pure-tone carrier. While the random-noise carrier exhibits random envelope fluctuations, the pure-tone carrier has a flat temporal envelope. In the latter case, the auditory system's resolution can only be limited by internal noise. The same is true for frozen-noise carriers that also do not exhibit any variability of stimulus parameters across presentation intervals but have, of course, a different distribution of power in the envelope spectrum than sinusoidal carriers. In this experiment, AM-detection thresholds are obtained for random- and frozen-noise carriers, as a function of the carrier bandwidth. The question is whether the detection performance in these conditions can be explained in terms of the "long-term" average distribution of the envelope power only, or whether the statistical variability of the envelope power is additionally needed to be taken into account.

### B. Method

#### 1. Subjects

Four male subjects participated in the experiment. Two subjects (SE and MB) also participated in the first experiment. All subjects had clinically normal hearing and had, with exception of subject CP, experience in other psychoacoustic experiments. Their age ranged from 27 to 29 years. Subjects MB, JD and CP were paid for their participation on an hourly basis.

#### 2. Apparatus, stimuli and procedure

The same apparatus, adaptive procedure and signal duration as in the first experiment were used. Only the differences are reported here. AM-detection thresholds were measured for modulation frequencies of 4, 16, 64, and 256-Hz, applied to either a random or deterministic ("frozen") noise carrier. The bandwidth of the carrier ranged from 8 to 2048 Hz, in two-octave steps (8, 32, 128, 512, 2048 Hz). The bandlimited noise carrier was generated in the temporal domain by selecting  $2^{17}$  samples ( $\approx 2.7$ s) from a Gaussian random variable. Bandlimiting was done in the frequency domain by zeroing the Fourier-coefficients outside the passband of the noise representation. In case of the random-noise carrier, this representation was refreshed prior to each trial and the three presentation intervals were consecutively cut from the 2.7s-representation. In case of the frozen-noise

carrier, a single noise representation was generated once prior to each threshold run and a fixed portion of that representation served as the carrier in all presentation intervals during the entire threshold run. Note, that this procedure results in a different frozen-noise realization for each subject and threshold run.

The equation describing the stimuli is:

$$s(t) = A [1 + m \sin(2\pi f_m t)] c(t), \quad (2.2)$$

where  $m$  is the modulation depth,  $f_m$  the modulation frequency and  $c(t)$  the carrier waveform. During a threshold run, the modulation depth  $m$  was varied in logarithmic steps ( $20 \log(m)$ ). The initial modulation depth was -6 dB. The initial step size was 4 dB and was divided by 2 after every second reversal until it reached the final stepsize of 1 dB. For each subject and threshold run, the threshold estimate was taken as the mean value of  $20 \log m$  across eight reversals obtained at the final stepsize. The thresholds reported are the mean of the estimates from three runs.

### C. Results and discussion

The mean data are shown in Fig. 2.4. The four panels show data for the modulation frequencies 4 Hz (upper left), 16 Hz (upper right), 64 Hz (lower left) and 256 Hz (lower right). In each panel, the data for the random-noise carrier are indicated by the circles while the squares represent the data for the frozen-noise carrier. Detection thresholds are plotted as a function of the carrier bandwidth. In all conditions, a threshold maximum is observed at a carrier bandwidth twice the modulation frequency. Except for the 4-Hz signal modulation, where the maximum occurs at the smallest bandwidth tested, thresholds decrease at both sides of the maximum. The thresholds obtained for the frozen-noise carrier are always lower than or equal to the thresholds obtained for the random-noise carrier. The data obtained for the frozen-noise carrier show a somewhat higher variability (indicated by the larger errorbars) than the random-noise data. This, however, is expected taking into account the relatively small number  $n$  of frozen-noise samples presented ( $n = 12$ , three for each of the four subjects). In contrast, when using the random-noise carrier, about 150 different random-noise samples are presented in each of the twelve runs. The maximum threshold difference between the two carrier types generally occurs at a carrier bandwidth near the maximum of the threshold pattern. The largest difference between the random and frozen-noise carrier can be observed for the lowest modulation frequency (4-Hz), when applied to the 8-Hz-wide carrier (upper left panel). Overall, the differences between the two carrier types decrease with increasing signal frequency. The maximum threshold difference for 16 and 64 Hz modulation frequency are 4 and 5 dB, respectively, in both cases for a carrier bandwidth of 128 Hz. For the highest modulation frequency (256 Hz), thresholds are very similar for the two carrier-noise type. A two-way ANOVA (MATLAB R12.1) using all conditions (carrier type [2] x carrier bandwidth [5]) showed a significant main effect of carrier type for signal-modulation frequencies of 4 and 16 Hz

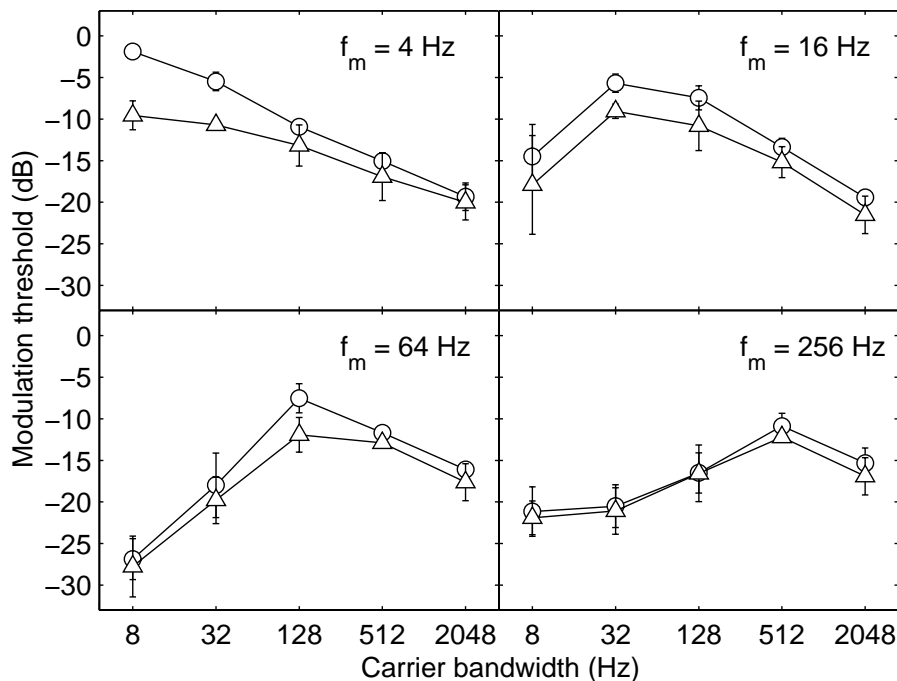


FIG. 2.4: Modulation-detection thresholds as a function of the bandwidth of a noise carrier. In each of the panels, the circles are for the random-noise carrier and the squares indicate data for the frozen-noise carrier. The different panels show average data for the modulation frequencies of 4 Hz (upper left), 16 Hz (upper right), 64 Hz (lower left) and 256 Hz (lower right).

( $p < 0.001$ ), and 64 Hz ( $p < 0.05$ ). No main effect of carrier type was found for the 256-Hz signal modulation.

AM-detection thresholds with random-noise carriers are mainly determined by the spectral distribution of the envelope fluctuations inherent to the noise carrier, as was already demonstrated in (Dau *et al.*, 1997a,b). Thresholds for a sinusoidal amplitude modulation, imposed on a bandlimited noise carrier, depend in a characteristic way on the carrier bandwidth. In the conditions considered here, the maximum threshold is observed when the carrier bandwidth equals two times the frequency of the signal modulation. It was demonstrated that such a threshold pattern can only be accounted for if a modulation-bandpass filter is assumed. A modulation-low-pass filter would fail in these conditions (Dau *et al.*, 1999).

The data in Fig. 2.4 demonstrate that the empirical threshold patterns obtained for the frozen-noise carriers are quite similar to those obtained for the random-noise carriers, i.e., listeners take only little advantage of the absence of any variability in the envelope waveform in case of the frozen-noise carrier. Thus, detection performance cannot be limited only by the stochastic nature of the envelope fluctuations inherent to the random-noise carrier but must be mainly affected by internal noise in the coding of the envelope as the remaining limiting factor in deterministic conditions.

Predictions obtained with the two models, EPSM and PEMO, are shown in Fig. 2.5 and Fig. 2.6 (closed symbols), respectively. Each figure also shows the corresponding empirical

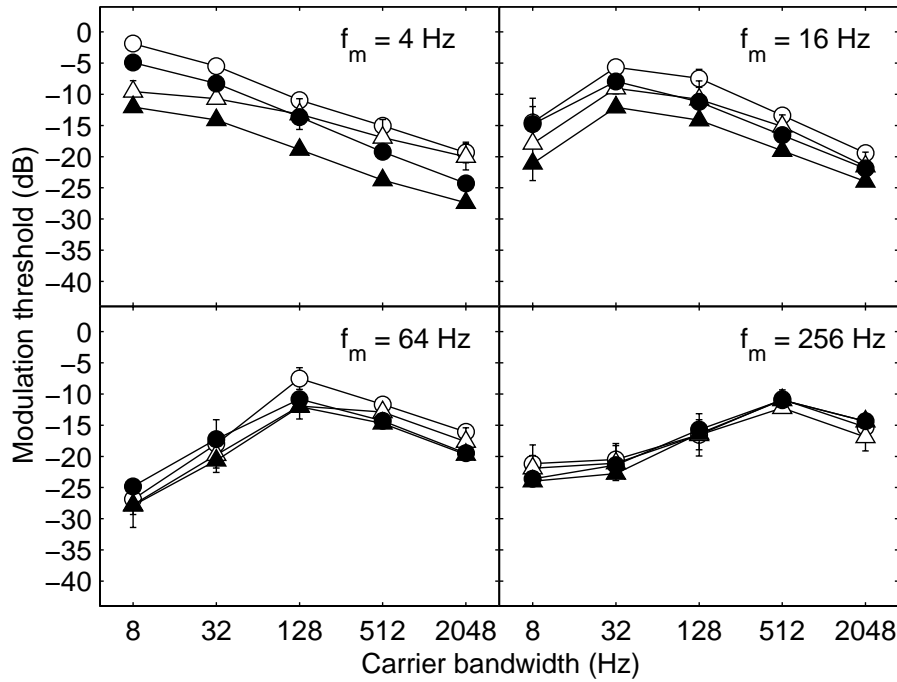


FIG. 2.5: Predictions obtained with EPSM (closed symbols) together with empirical data (open symbols), replotted from Fig. 2.4. The four panels show modulation-detection thresholds for signal modulations of 4, 16, 64, and 256-Hz, respectively, as a function of the bandwidth of the carrier. The circles represent thresholds for the random-noise carrier and the upward-pointing triangles show results for the frozen-noise carrier.

data (open symbols) from Fig. 2.4. The four panels show AM-detection thresholds for the signal frequencies 4, 16, 64, and 256-Hz, as a function of the carrier bandwidth. The circles represent thresholds for the random-noise carrier and the triangles show results for the frozen-noise carrier.

The EPSM (Fig. 2.5) captures all main effects apparent in the data. It accounts for the general shape of the threshold patterns as well as for the differences between random and frozen noise. As observed in the data, these differences decrease with increasing signal frequency. Within EPSM, the reason for the diminishing difference is that with increasing signal-modulation frequency (and thus increasing bandwidth of the constant-Q modulation filter), the standard deviation of the integrated envelope power decreases relative to the mean value, as more and more components of the envelope spectrum fall within the passband of the filter. For 256 Hz signal modulation, the mean integrated envelope power becomes the limiting quantity in both cases, the random- and the frozen-noise carrier. The other model, PEMO (Fig. 2.6), also accounts for the data in the random-noise carrier conditions (circles). Here, the model produces similar predictions as EPSM. However, PEMO completely fails in all frozen-noise carrier conditions (filled triangles) where it predicts a constant threshold of about -35 dB, independent of carrier bandwidth and signal frequency. This is a direct consequence of the constant-variance internal noise that does not depend on the magnitude of the mean envelope power of the stimulus.



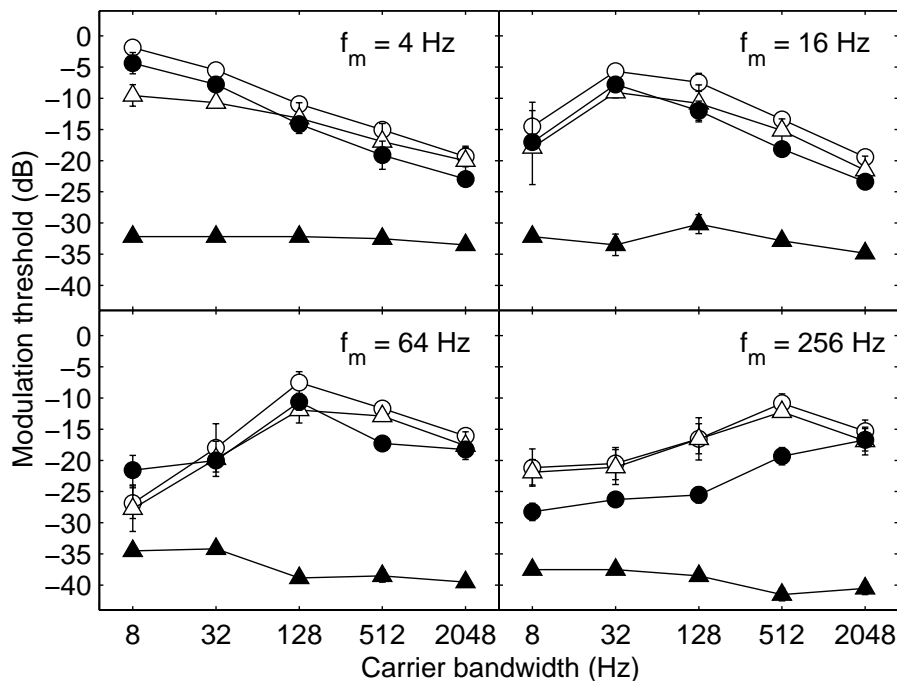


FIG. 2.6: Predictions obtained with PEMO (closed symbols) together with the empirical data (open symbols) as in Fig. 2.5.

## V. AM MASKING AS A FUNCTION OF MASKER VARIABILITY

### A. Rationale

In the previous experiment, the detection performance for the signal modulation was limited by the inherent envelope fluctuations of the carrier noise. The current experiment examines modulation masking as a function of the statistical properties of a narrowband-noise masker modulation, centered at the signal-modulation frequency. The experiment is designed to allow for a more gradual transition between deterministic and stochastic stimuli: semi-frozen and rms-equalized random-noise maskers are used in addition to frozen- and random-noise maskers. This is done in order to examine the validity of a “long-term” average quantity as the observation variable, such as the mean envelope power within the framework of EPSM, and in order to investigate the role of memory in experiments with frozen and semi-frozen stimuli.

### B. Method

#### 1. Subjects

Four male subjects participated in the experiment. Three of the subjects (MB, JD and author SE) also participated in the second experiment. Except the fourth subject (PE), all subjects had experience in psychoacoustic experiments. All subjects had clinically normal hearing. Their age ranged from 25 to 36 years. Subjects MB and PE were paid on an

hourly basis for their services.

## 2. Apparatus, stimuli and procedure

The same apparatus, adaptive procedure and signal duration were used as in the previous experiment. Detection thresholds for 4, 16 and 64-Hz signal modulation were measured in the presence of a narrowband-noise masker modulation. Signal and masker modulation were applied to a 4-kHz pure-tone carrier. The masker was geometrically centered at the signal-modulation frequency and had a half-octave bandwidth (1.4, 5.6 and 22.4 Hz). The equation describing the stimuli was:

$$s(t) = A [1 + m \sin(2\pi f_m t) + n_m(t)] c(t), \quad (2.3)$$

where  $m$  is the signal-modulation depth,  $f_m$  the signal-modulation frequency, and  $c(t)$  denotes the carrier waveform. The masker modulation  $n_m(t)$  was either frozen noise, semi-frozen noise, rms-equalized random noise, or random noise. All noise-masker modulations were generated in the temporal domain by selecting  $2^{17}$  samples ( $\approx 2.7$ s) from a Gaussian random variable. The signals were then transformed to the frequency domain, bandlimited by zeroing the Fourier-coefficients outside the desired passband, and transformed to the time domain by inverse Fourier transformation. For the frozen-noise masker, the identical noise realization was used in each interval. The rms value of this masker was set to -15 dB. In case of the semi-frozen noise, a different frozen-noise realization was used for each trial during a threshold run, i.e., the masker modulation was only frozen throughout the three presentation intervals but not during the entire run. In case of the two random maskers, the 2.7-s long representation was refreshed prior to each trial and three different portions were cut out randomly for each presentation interval. For the rms-equalized random noise, the rms value of each masker realization was set to -15 dB, while for the random-noise masker, the rms value of the entire 2.7-s representation was set to -15 dB. In the latter case, only the long-term average across the rms values of the individual masker waveforms was -15 dB.

## C. Results and discussion

Figure 2.7 shows the mean detection thresholds, averaged across the subjects, for the signal-modulation frequencies 4 Hz (upper left panel), 16 Hz (upper right panel) and 64 Hz (lower left panel). The four different masking-noise conditions are indicated on the abscissa: frozen noise (Fro), semi-frozen noise (SFro), rms-equalized random noise (EqRan), and random noise (Ran). For each signal frequency, the lowest detection threshold is observed for the frozen-noise masker. For the other three noise-masker types, thresholds are increased by about the same amount. With increasing signal-modulation frequency, the detection threshold decreases in all masking conditions. The threshold difference between 4 Hz and 64 Hz amounts 4 dB, averaged across all conditions. The maximum difference is observed for EqRan noise (6 dB), while the minimum difference (2 dB) is observed in the Fro condition.

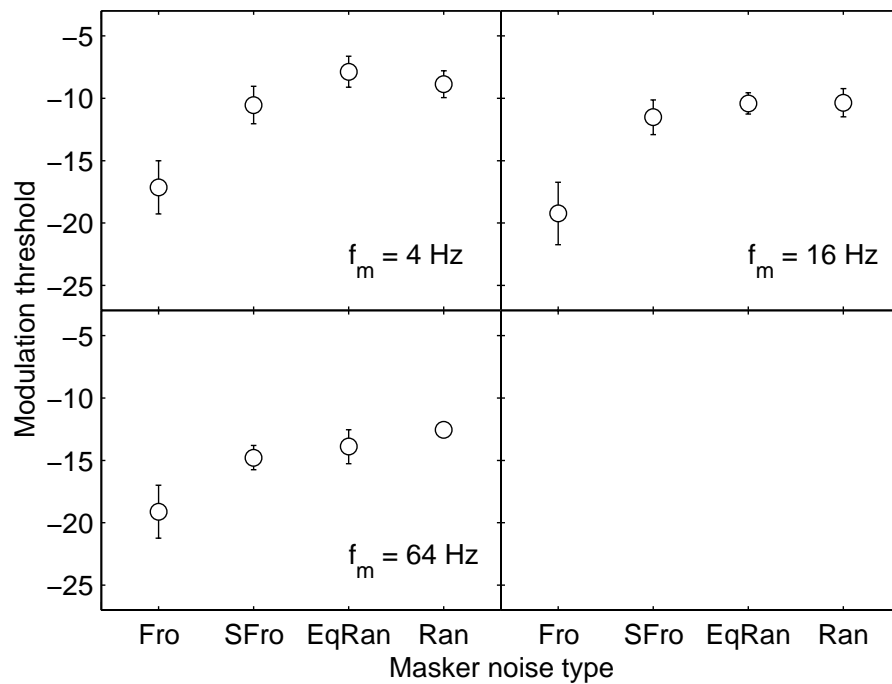


FIG. 2.7: Masked-modulation-detection thresholds in the presence of different types of a narrowband (half-octave) noise. The panels display the data for the different signal-modulation frequencies 4 Hz (upper left), 16 Hz (upper right) and 64 Hz (lower left). The noise masker was always centered at the signal-modulation frequency. The abbreviations at the abscissa indicate frozen (Fro), semi-frozen (SFro), equalized random (EqRan), and random (Ran) noise maskers.

The effect of carrier-noise type was investigated using a one-way ANOVA (MATLAB R12.1). A highly significant main effect of noise type was found for all signal frequencies ( $p < 0.001$ ). Post hoc comparisons based on Tukey's honestly significant difference (HST) criterion showed that the threshold in the frozen-noise condition was significantly different from all other conditions ( $p < 0.001$  for 4 and 16 Hz,  $p < 0.01$  for 64 Hz). No difference between the other conditions was found. Thus, for all signal frequencies tested, the listener's detection performance significantly improved only in the case of the frozen-noise masker. Although the other three types of maskers differed in their variability, all three resulted in similar masked thresholds. The large threshold difference between the frozen and the semi-frozen masker conditions is somewhat surprising. Decisions seem to be more based on the comparison of an internally formed expectation for the target signal with each of the signals in the intervals of a trial, rather than on the comparison of the three intervals without any history. Thus, memory effects during the threshold run seem to play an important role. Unlike it might have been expected from the deterministic nature of each of the trials in the semi-frozen condition, listeners are not able to improve their detection performance relative to the two stochastic conditions (equalized random and random). It is possible, that the significantly lower performance in the semi-frozen condition compared to the frozen condition was observed because the different masker types were presented in a randomly mixed order. While a frozen-noise run can be easily

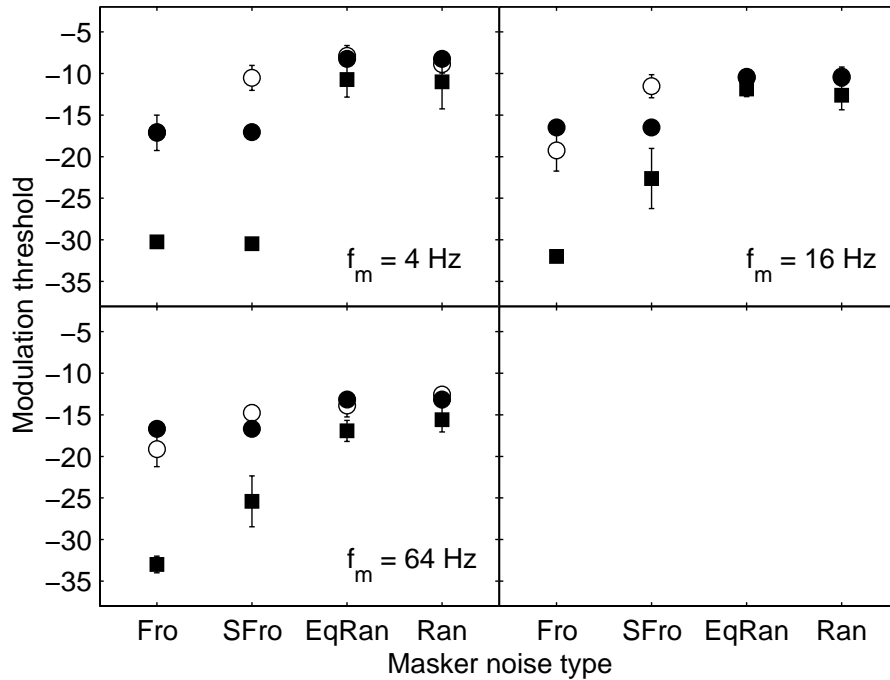


FIG. 2.8: Comparison of model predictions for EPSM (close circles) and PEMO (closed squares) with the empirical data from Fig. 2.7 (open symbols). The abbreviations at the bottom represent the different masker types, frozen (Fro), semi-frozen (SFro), equalized random (EqRan), and random (Ran).

identified out of the three others, listeners may not be able to identify a semi-frozen run. It can only be speculated whether prior knowledge of the masker type would lead to a better performance in the semi-frozen runs.

Figure 2.8 shows the predictions of the two models (filled symbols) in these conditions. The experimental data are replotted and indicated as open symbols. As mentioned in Section II., the two models do not only differ in their assumptions about the internal variability, inherent in the coding process, but also differ in terms of considering memory effects in their decision devices. Both models account for the thresholds in the rms-equalized and the random-noise conditions. However, neither of the two models is able to capture all aspects of the data. EPSM (closed circles) predicts a threshold difference between the random and frozen conditions. However, it cannot, per definition, distinguish between the two random conditions, and between the two frozen conditions. In the two random conditions, the observation variable within EPSM, the long-term average integrated envelope power, is the same. Considering EPSM as a memoryless model, it is clear that it cannot predict any difference between the frozen and semi-frozen condition. Since all stimuli in a trial are deterministic in the frozen and the semi-frozen condition, detection performance within EPSM is purely limited by internal noise. As in the previous experiment, PEMO largely overestimates detection performance in the two frozen-masker conditions. However, except for the 4-Hz condition, this model predicts a threshold difference between the frozen and semi-frozen condition. The reason for this difference is that PEMO, in contrast to EPSM, includes memory effects in the detection process. The model permanently “recalls” specific

features of the internal representation of the stimulus (the template), derived by averaging a supra-threshold representation of the signal. The template is “perfect” for deterministic stimuli while it is corrupted by the external variability in case of stochastic stimuli. It can be assumed that the template mechanism quite realistically simulates the formation and short-term presence of a “matched” feature-selective filter during a threshold run. Despite the fact that the overall deviation from the data in the frozen and semi-frozen condition is larger for PEMO than for EPSM, PEMO can, in principle, account for a difference in performance between the frozen and the semi-frozen condition. The relatively larger overall deviation is, again, a consequence of the constant-variance internal noise.

## VI. OVERALL DISCUSSION

AM-depth discrimination with broadband noise and pure-tone carriers (first experiment) showed that Weber’s law holds for standard modulation depths well above absolute (modulation) detection threshold. It does not hold for small standard depths where a constantly rising Weber fraction is observed, as has already been shown in Wakefield and Viemeister (1990) for broadband-noise carriers. The present study demonstrates that, in the framework of AM-processing models, the assumption of a “Weber-fraction”-type internal noise is essential in order to account for the data. The fact that Weber’s law does not apply over the entire range of standard depths for the noise carrier can be explained by the external variability inherent to the envelope of the noise carrier. The envelope power spectrum model, EPSM, is able to account for the data by combining internal noise that is proportional to the mean of the observation variable, the integrated envelope power, and external noise as inherent property of the stimuli. The perception model, PEMO, accounts very well for the data as long as the stimulus variability limits performance, while it cannot predict Weber’s law since a constant-variance internal noise is assumed, that is independent of the mean of the observation variable.

The empirical findings of AM detection with random- and frozen-noise carriers (second experiment) are perfectly in line with the assumption that a Weber-fraction-type internal noise in combination with the external variability of the stimuli limits detection performance. The comparison of the data in the random- and frozen-noise conditions clearly demonstrates that listeners gain only little from the absence of external stimulus variability as given in the frozen-noise conditions. This proves, on the one hand, that internal noise plays a dominant role in conditions with frozen-noise carriers. It also demonstrates, on the other hand, that the internal noise must be proportional to the envelope power of the stimuli (at the output of an envelope-frequency selective process), since the data for random- and frozen-noise carriers follow the same threshold pattern. A different behaviour would be expected from an “ideal” observer as reflected by the threshold predictions of PEMO. In the absence of external variability, thresholds are only determined by a constant amount of internal noise, independent of the stimulus. The EPSM behaves more like a “real” observer, severely hampered by internal noise that is proportional to the mean of

the observation variable.

The role of internal and external limitations in AM detection were further investigated in the third experiment, where thresholds were obtained in the presence of a narrowband-noise masker as a function of the degree of the variability in the masker. The most interesting observation was that masking is considerably reduced in the frozen condition compared to the semi-frozen condition. This result could not have been expected as long as it is assumed that decisions within the alternative forced choice task are based on a comparison of the stimuli within a trial. In this case, there should be no difference between the frozen and semi-frozen condition. However, the listeners seem to have developed a different strategy in the frozen than in the semi-frozen condition. This result draws some interesting conclusions for the two models of the present study. Within EPSM, decisions are based on the increase of the observation variable in comparison to the standard intervals only. This results in the same predictions for the frozen and the semi-frozen masker noise. The fact that the observation variable is a long-term average quantity, leads to the same thresholds predictions for the two random-noise conditions. The latter observation is in line with the data. The other model, PEMO, while overestimating detection performance in the two frozen-noise conditions, is in principle able to predict a threshold difference between the frozen and the semi-frozen condition. This can be explained in terms of the increased amount of information reflected in the observation variable within PEMO, in comparison to EPSM, independent of the above described differences in the internal noise process. PEMO calculates the cross-correlation coefficient (cross-power) between the current stimulus representation and a stored template representation which incorporates internal memory and information about the temporal course of the internal representation. In contrast, EPSM only uses a strongly reduced amount of information: The envelope power spectrum represents the Fourier transform of the auto-correlation of the internal stimulus representation, neglecting memory and temporal information.

Taken together, both models successfully describe the data when the external variability of the stimuli limits performance in AM-depth discrimination and AM detection. The most severe difference between the two models in these conditions is reflected in the assumptions about the nature of the internal-noise process. As a consequence, one model (EPSM) accounts for the data in deterministic conditions while the other model (PEMO) fails in these conditions. Another difference between the two models is the observation variable. The observation variable in PEMO, based on a cross-correlation, might be more realistic than the integrated envelope power in EPSM. Future models of AM detection should probably combine the more realistic internal variability, as assumed in EPSM, with the more realistic observation variable as assumed in PEMO. Future versions of PEMO will have to include Weber's law in the envelope domain. With such a modification, this model should also be able to account for deterministic stimulus conditions while all predictions with random stimuli should remain unchanged.

## VII. CONCLUSIONS

This study demonstrated that effects of (external) stimulus envelope variability and effects of internal envelope coding in the auditory system can be separated as two independent limiting factors in AM-depth discrimination and AM-detection experiments. The following conclusions can be drawn from the present study:

- AM-depth discrimination thresholds for a modulation frequency of 16 Hz closely follow Weber’s law for large standard AM depths, when applied to a broadband-noise or a pure-tone carrier. Discrimination thresholds show a systematic deviation from Weber’s law for standard depths below -8 dB in case of the broadband-noise carrier, and below -18 dB in case of the pure-tone carrier. In this region, a constant absolute AM-depth increment is required for detection rather than a constant relative increment.
- AM-detection thresholds with frozen- and random-noise carriers follow a similar characteristic pattern as a function of the carrier bandwidth. The maximum masked threshold is observed when the carrier bandwidth equals twice the signal-modulation frequency which can be explained as a result of an envelope-frequency selective process. In general, modulation-detection thresholds are lower for the frozen- than for the random-noise carrier. The maximum threshold difference between the frozen and random condition is observed for signal-modulation frequencies  $\leq 16$  Hz. Differences diminish with increasing carrier bandwidth and increasing signal-modulation frequency.
- The empirical findings can be accounted for by the envelope power spectrum model (EPSM) that assumes a constant relative detection criterion and takes the stimulus envelope variability into account. The perception model (PEMO) accounts for the data as long as the envelope variability limits the detection performance. Since PEMO assumes a constant absolute detection criterion in the envelope domain, the model fails to account for the data in deterministic conditions where internal noise inherent to the internal coding mechanisms limits the detection performance.
- Masked AM thresholds in the presence of a narrowband-noise masker modulation are considerably lower for a frozen-noise masker than for a semi-frozen- or random-noise masker. It is shown that listeners are not able to gain information from identical masker-modulation waveforms within the trials (semi-frozen) of a 3-alternative forced choice (3-AFC) run, while the gain is significant when the identical masker waveform is presented during the entire threshold run (frozen). This observation contradicts the assumption that decisions in a 3-AFC task are based on a “memoryless” comparison of the three intervals within a single trial. As a consequence, models without memory (like the EPSM) fail to correctly mimic the decision process in human listeners.

## **ACKNOWLEDGMENTS**

We thank Stefan Uppenkamp for comments on an earlier version of the manuscript. We are grateful to Birger Kollmeier for his support. This work was supported by the Deutsche Forschungsgemeinschaft (DFG), Research Project SPP 1046.



## CHAPTER 3

# SPECTRO-TEMPORAL PROCESSING IN THE ENVELOPE-FREQUENCY DOMAIN<sup>a</sup>

### ABSTRACT

The frequency selectivity for amplitude modulation applied to tonal carriers and the role of beats between modulators in modulation masking were studied. Beats between the masker and signal modulation as well as intrinsic envelope fluctuations of narrowband-noise modulators are characterized by fluctuations in the “second-order” envelope (referred to as the “venelope” in the following). In experiment 1, masked-threshold patterns (MTPs), representing signal modulation threshold as a function of masker-modulation frequency, were obtained for signal-modulation frequencies of 4, 16, and 64 Hz in the presence of a narrowband-noise masker modulation, both applied to the same sinusoidal carrier. Carrier frequencies of 1.4, 2.8, and 5.5 kHz were used. The shape and relative bandwidth of the MTPs were found to be independent of the signal-modulation frequency and the carrier frequency. Experiment 2 investigated the extent to which the detection of beats between signal and masker modulation is involved in tone-in-noise (TN), noise-in-tone (NT), and tone-in-tone (TT) modulation masking, whereby the TN condition was similar to the one used in the first experiment. A signal-modulation frequency of 64 Hz, applied to a 2.8-kHz carrier, was tested. Thresholds in the NT condition were always lower than in the TN condition, analogous to the masking effects known from corresponding experiments in the audio-frequency domain. TT masking conditions generally produced the lowest thresholds and were strongly influenced by the detection of beats between the signal and the masker

---

<sup>a</sup>This chapter is published as S. D. Ewert, J. L. Verhey, and T. Dau, (2002). “Spectro-temporal processing in the envelope-frequency domain”, *J. Acoust. Soc. Am.* **112**, 2921-2931.

modulation. In experiment 3, TT masked threshold patterns were obtained in the presence of an additional sinusoidal masker at the beat frequency. Signal-modulation frequencies of 32, 64 and 128 Hz, applied to a 2.8-kHz carrier, were used. It was found that the presence of an additional modulation at the beat frequency hampered the subject's ability to detect the envelope beats and raised thresholds up to a level comparable to that found in the TN condition. The results of the current study suggest that (i) envelope fluctuations play a similar role in modulation masking as envelope fluctuations do in spectral masking, and (ii) envelope and envelope fluctuations are processed by a common mechanism. To interpret the empirical findings, a general model structure for the processing of envelope and envelope fluctuations is proposed.

## I. INTRODUCTION

The temporal envelope of a stimulus is known to play an important role for the perception of everyday-life sounds. Even in psychoacoustic experiments that are primarily related to the *spectral* resolution of the auditory system, such as masking patterns, the information of the temporal envelope can often not be ignored. Masking patterns have long been used to illustrate the frequency selectivity of the auditory system and have been used as a tool for estimating the spread of excitation of the masker within the cochlea (e.g., Wegel and Lane, 1924; Fletcher and Munson, 1937; Zwicker, 1956). They are typically obtained by measuring the threshold for a signal as a function of the signal frequency in the presence of a masker with fixed frequency and level. The masking patterns of narrow-band sounds, either sinusoids or bands of noise, have been measured in many experiments. It became apparent that the masking patterns showed complex features that could not be readily explained in terms of spread of excitation. In particular, temporal envelope fluctuations (modulations) were shown to strongly influence the "spectral masking" data in some conditions. For example, if the masking of a tonal signal by a tonal masker (tone-in-tone masking) is measured, the beats between the masker and the signal provide an important cue if the spectral separation between the signal and the masker is small (e.g., Wegel and Lane, 1924; Fletcher and Munson, 1937; Egan and Hake, 1950; Ehmer, 1959; Moore *et al.*, 1998). Interestingly, beats appear to be less salient if a narrowband noise is used instead of a tone as the masker (tone-in-noise masking). In such a condition, it is assumed that the intrinsic envelope fluctuations of the noise at least partly mask the beat cue (e.g., Moore *et al.*, 1998). On the other hand, these inherent fluctuations of the noise can also provide a detection cue if the role of the signal and the masker is reversed, i.e., if a narrowband noise acts as the signal and is masked by a tone (noise-in-tone masking). In such a reversed condition, much lower thresholds are generally observed (Moore *et al.*, 1998; Hall, 1997).

The role of temporal cues, such as beats and intrinsic envelope fluctuations, in spectral masking has recently been investigated quantitatively by Derleth and Dau (2000) and Verhey (2002). Predictions from the modulation-filterbank model by Dau *et al.* (1997), which

performs a combined spectro-temporal signal analysis, were compared to experimental data from Moore *et al.* (1998) and Hall (1997). The model could generally describe the masking data reasonably well, whereas predictions from a version of the model that acts like an energy detector failed in some conditions. Within the framework of the processing model, the modulation filterbank concept allowed for the explanation of both the detection of beats as well as the masking of beats when inherent envelope fluctuations of similar spectral content as the beats are present at the same time.

Several recent studies of amplitude-modulation detection with two- or multi-component modulators have revealed that beats between the components may become a salient cue in modulation masking experiments (e.g., Fassel, 1994; Strickland and Viemeister, 1996; Sheft and Yost, 1997; Ewert and Dau, 2000). Furthermore, Moore *et al.* (1999) demonstrated that the threshold for 5-Hz signal modulation was affected by the presence of a two-component masker modulator beating at a 5-Hz rate. The threshold was dependent on the phase of the signal modulation relative to the beat cycle of the masker modulators. In two recent studies, Lorenzi *et al.* (2001a,b) measured detection thresholds for slow sinusoidal variations of the modulation depth of a “carrier” amplitude modulation applied to noise and pure-tone carriers. It was shown that in some conditions the resulting “second-order” modulation detection thresholds were close to those for “first-order” modulations. The above findings suggest that these slow fluctuations, not present in the envelope spectrum of the stimuli, are perceptually relevant in several experimental conditions.

Second-order modulations and beats between modulators can be mathematically well described in terms of the second-order envelope for arbitrary stimuli (see Appendix B). Throughout the present paper, the new term “venelope” is used to describe the envelope of the ac-coupled envelope. Figure 3.1 illustrates the relation between envelope and venelope. The upper left panel shows the temporal waveform of a sinusoidal carrier that is modulated at two frequencies (64 and 80 Hz). The lower left panel shows the envelope of the stimulus (solid line). The envelope spectrum is given in the upper right panel of the figure. While the envelope spectrum shows two peaks at the frequencies of the primary components, the temporal waveform also contains a slow variation of the modulation depth, with a periodicity equal to the difference frequency of the two primary components. The venelope is indicated as the dashed line in the lower left panel of Fig. 3.1. The lower right panel shows the spectrum of the venelope. In the given example, the venelope is a rectified sinusoid and the venelope spectrum consists of equally spaced components with a fundamental at the beat frequency and harmonics that decrease in amplitude with increasing frequency.

The present study investigates whether the venelope can produce similar effects in masking experiments in the envelope-frequency domain as were found for the envelope in the audio-frequency domain. In analogy to the spectral masking experiments in the audio-frequency domain, tones and narrowband noises were used as the signal and the masker modulation. In order to avoid the possibility that inherent fluctuations of the carrier influence the results, sinusoidal carriers were used in the present study. Experiment 1

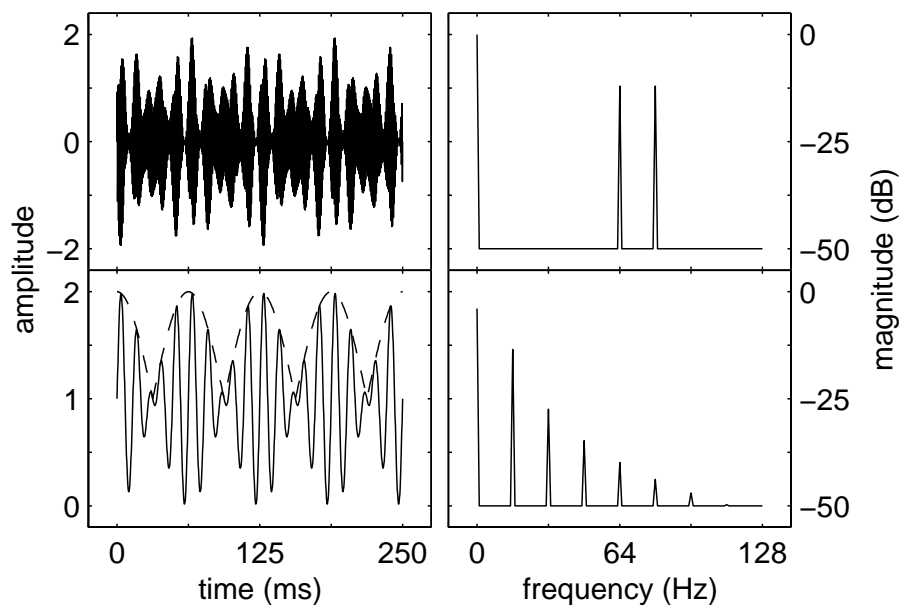


FIG. 3.1: Sample representation of the stimuli used in the present study. The upper left panel shows the waveform of a 2.8-kHz tone modulated by the sum of two sinusoids at 64 and 80 Hz, respectively. Its envelope (solid curve) and venelope (dashed curve) are shown in the lower left panel. The upper right panel displays the envelope-power spectrum of the stimulus. The lower right panel shows the venelope-power spectrum.

investigates modulation masking using carrier frequencies of 1.4, 2.8, 5.5 kHz. These frequencies were chosen to cover the mid-frequency audio range from earlier modulation-masking studies with broadband noise carriers (Bacon and Grantham, 1989; Houtgast, 1989; Ewert and Dau, 2000). The goal was to obtain information on modulation processing in the three separate frequency regions. The modulation masking paradigm was the same as described earlier in Ewert and Dau (2000): A masked threshold pattern (MTP) was obtained for a fixed signal-modulation frequency in the presence of a narrowband-noise masker modulation. The spectral position of the masker modulation was shifted relative to the signal modulation. The experiment served as a reference for experiments 2 and 3, described below. The fixed signal frequency approach was also used by Takahashi and Bacon (1992) and Lorenzi (1997). This paradigm is particularly suitable for the derivation of modulation filters. In contrast, masking patterns where a fixed masker and a variable signal frequency is used (e.g., Houtgast, 1989; Bacon and Grantham, 1989), would reflect a “modulation excitation pattern” evoked by the masker modulation, analogous to the excitation patterns commonly described in the audio-frequency domain (e.g., Zwicker and Feldtkeller, 1967; Moore and Glasberg, 1986). Hence, the shape of the masking pattern for a particular masker would not directly reflect the shape of a specific modulation filter.

In Experiment 2, masking patterns with different stimulus types for the signal and the masker are investigated. Narrowband noises or tones were used for the signal and the masker modulation and several signal-masker combinations were considered. The experiment is focused on a set of parameters where beats between the signal and masker

modulation might have played a role in Experiment 1. The experiment examines whether beats between modulations play the same role in tone-in-noise, noise-in-tone and tone-in-tone masking in the envelope-frequency domain, as they do in comparable masking experiments in the audio-frequency domain (e.g., Moore *et al.*, 1998).

Experiment 3 addresses the question whether envelope fluctuations and envelope fluctuations (the beats between the modulator components) can interact when they lie in a similar range of frequencies. A masking effect might be expected from the results of Moore *et al.* (1999), who showed an interference when the task was to detect an amplitude modulation in the presence of a envelope component at the same frequency. However, Lorenzi *et al.* (2001b) showed that the detection of envelope beat cues cannot be abolished when using a narrowband noise carrier. In this case, the inherent envelope fluctuations of the carrier showed only a small effect on the detectability of envelope fluctuations in the same frequency range. Experiment 3 investigates whether the detection of beats between a sinusoidal signal modulation and a sinusoidal masker modulation is influenced by an additional sinusoidal masker modulation at the beat frequency.

Based on the experimental findings, a general model structure for the processing of the envelope and the envelope is proposed.

## II. EXPERIMENT 1: ENVELOPE-FREQUENCY SELECTIVITY USING SINUSOIDAL CARRIERS

### A. Method

#### 1. Subjects

Three normal-hearing subjects participated in the study. Their age ranged from 26 to 29 years. All subjects had experience in other psychoacoustic experiments. One of the subjects (S3) was author SE, the other subjects were paid an hourly wage for their participation.

#### 2. Apparatus and stimuli

Subjects listened diotically via Sennheiser HD 25 headphones while seated in a sound-attenuating booth. Signal generation and presentation during the experiments were computer controlled using the signal-processing software package SI developed at the Drittes Physikalisches Institut at the University of Göttingen. The stimuli were digitally generated on a Silicon Graphics workstation at a sampling rate of 32 kHz and converted to analog signals by an onboard two-channel 16-bit DAC including reconstruction filtering. A sinusoidal signal modulation and a narrowband Gaussian-noise masker modulation were applied to a pure-tone carrier. Sinusoidal carriers at 1.4 kHz, 2.8 kHz, and 5.5 kHz were used. The duration of the carrier was 600 ms including 50-ms  $\cos^2$  onset and offset ramps. The modulation started 50 ms after carrier onset, ended 50 ms before carrier offset, and

was gated with 50-ms  $\cos^2$  ramps. The signal modulation frequency was 4, 16 or 64 Hz. For each signal frequency, the spectral position of the masker modulation was varied in the range from -2 to +2 octaves relative to the signal frequency, using a step size of  $2/3$  octaves. In order to avoid spectral cues because of partly resolved spectral components, the maximum masker-modulation frequency was chosen in proportion to the carrier frequency. It was 64, 128, and 256 Hz for the carrier frequencies 1.4, 2.8, and 5.5 kHz, respectively. The bandwidth of the masker modulation was fixed at 1.4, 5.6, and 22.3 Hz for the signal frequencies of 4, 16, and 64 Hz, respectively. For the masker centered at the signal frequency (on-frequency condition) this corresponded to a  $1/2$  octave bandwidth. The root-mean-square (rms) modulation depth of the masker was  $-10$  dB in all conditions. The signal modulation started at a positive going zero-crossing. A masker realization of  $2^{16}$  samples (approximately 2 seconds) was generated before each threshold run. In each presentation interval during the experimental run, the masker waveform  $m_m(t)$  was cut randomly from the long realization. The stimuli are as follows:

$$s(t) = a \{ \sin(2\pi f_c t) [1 + m \sin(2\pi f_s t)] [1 + b m_m(t)] \}, \quad (3.1)$$

where  $f_c$  is the carrier frequency,  $m$  is the signal modulation depth and  $f_s$  represents the signal-modulation frequency. Setting  $b$  to zero eliminates the masker modulation, as used to determine the (unmasked) reference threshold. This multiplicative approach of combining signal and masker modulation was also used in Houtgast (1989) and Ewert and Dau (2000). To avoid possible level cues due to the presence of the signal modulation, the stimuli were adjusted to have equal energy in each interval of the forced-choice trial. The overall presentation level was 65 dB SPL.

### 3. Procedure

A three-interval, three-alternative forced-choice paradigm was used to measure modulation-detection thresholds. The psychophysical task was to identify the one randomly chosen interval containing the signal modulation. The two other intervals contained either no modulation or only the masker modulation. The modulation depth of the signal was varied in dB ( $20 \log m$ ) using a one-up two-down procedure, estimating the 70.7 % correct point of the psychometric function (Levitt, 1971). The three observation intervals were separated by 500 ms. The step size in each run was initially 4 dB and was divided by 2 after every second reversal until it reached 1 dB. At this step size, 8 reversals were obtained and the threshold estimate was calculated as the mean value of  $20 \log m$  at these reversals. Each threshold reported represents the mean of the estimates from at least three runs. On the rare occasions when the standard deviation of the three estimates exceeded 3 dB, an additional estimate was obtained and the first estimate was discarded.

## B. Results

The pattern of results was similar for the three subjects, so the mean data and standard deviations are shown in Fig. 3.2. The three panels show data for the different carrier

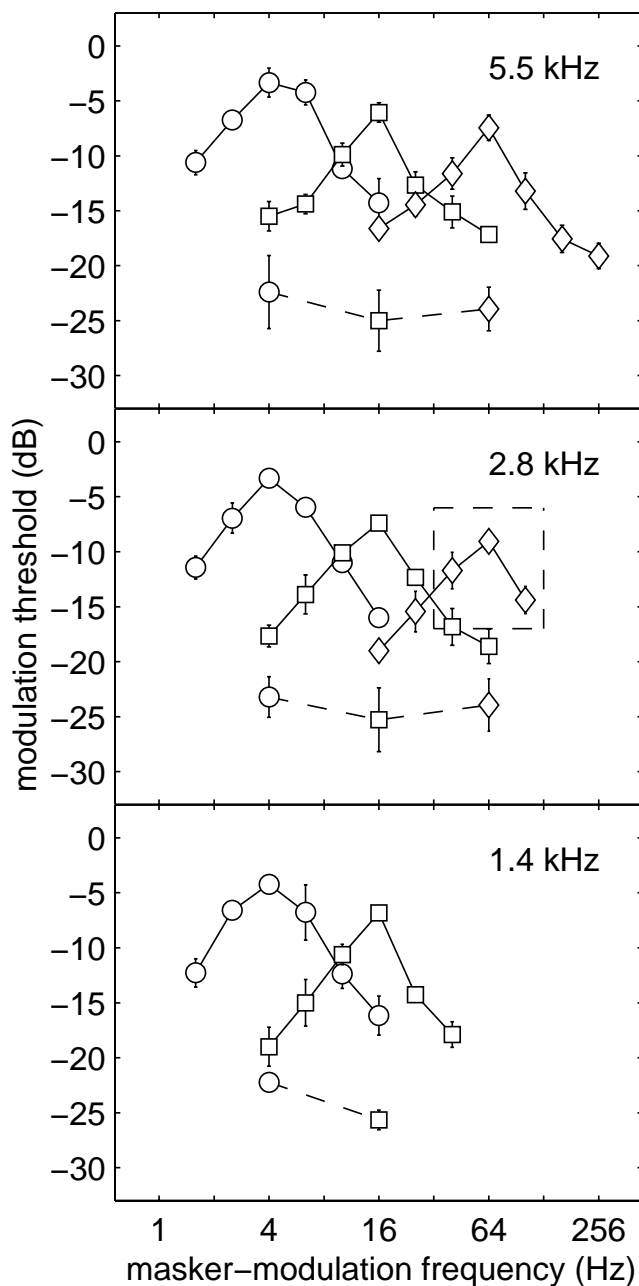


FIG. 3.2: Average masked-threshold patterns for signal frequencies of 4 Hz (circles), 16 Hz (squares), 64 Hz (diamonds) and carrier frequencies of 5.5 kHz (upper panel), 2.8 kHz (middle panel), and 1.4 kHz (bottom panel). The signal-modulation depth at threshold is plotted as a function of the masker frequency. The masker level was always -10 dB. In each on-frequency condition, the noise masker was half-octave wide. The absolute bandwidth was held constant when the masker was shifted in the range from -2 to +2 octaves relative to the signal frequency (see text). The dashed curve represents the reference thresholds without modulation masker. The dashed box refers to the focus of the second experiment.

frequencies of 5.5 kHz (top panel), 2.8 kHz (middle panel), and 1.4 kHz (bottom panel). Masked thresholds (solid lines) are shown for the signal frequencies 4 Hz (circles), 16 Hz (squares), and 64 Hz (diamonds). In addition, each panel shows the “absolute” thresholds where no masker modulation was presented (dashed line). For the different signal frequencies, these thresholds are indicated by the corresponding symbols connected by the dashed line. Note that for the lowest carrier frequency (1.4 kHz, bottom panel) the 64-Hz signal pattern was not obtained since spectrally resolved sidebands most likely would have influenced the results. The results are very similar across carrier frequency. All patterns show a peak when the masker center frequency was equal to the signal frequency. In general, the patterns are roughly symmetric and invariant in width on a relative (logarithmic) frequency scale.

TABLE 3.1: Best fitting Q-values for the second-order bandpass filters at center frequency  $cf$  as assumed within the framework of the EPSM. The value was fitted to the masked-threshold patterns from Fig. 3.2. A least-square fitting procedure was used. Corresponding  $-3$ -dB bandwidths of the filters are given in parentheses. The bar denotes a condition where no empirical data were obtained.

$cf$ (Hz)	carrier frequency		
	1.4 kHz	2.8 kHz	5.5 kHz
4	1.29 (3.1)	1.24 (3.2)	1.10 (3.6)
16	1.82 (8.8)	1.16 (13.8)	1.03 (15.5)
64	-	1.15 (55.5)	1.25 (51.4)

The patterns for the 4-Hz signal seem to be slightly broader compared to the patterns for the higher frequencies. The 4-Hz patterns also exhibit the highest overall thresholds, with a maximum threshold of  $-3.6$  dB, averaged across all carrier frequencies. For 16 Hz and 64 Hz the maximum average threshold is at  $-6.8$  dB and  $-8.2$  dB, respectively.

In order to obtain a more quantitative estimate of envelope-frequency selectivity, the transfer function of a second-order bandpass filter combined with a first-order lowpass filter was fitted to the empirical data. The same method and underlying envelope power spectrum model (EPSM) as proposed in Ewert and Dau (2000) were used. This model is conceptually related to the power spectrum model of masking (Fletcher, 1940; Patterson and Moore, 1986) in the audio-frequency domain. It is assumed that thresholds are related to the integrated envelope power in the passband of a modulation filter centered at the signal-modulation frequency. A second-order bandpass filter was chosen for the modulation filter. The bandwidth of the filter is easily described with a single variable, the Q-value. Its transfer function provides the required symmetry on a logarithmic frequency scale, as observed in the data. In order to account for the increasing asymmetry of the filter shape observed for the higher masker center frequencies ( $> 64$ ), a first-order lowpass filter with a cutoff frequency of 150 Hz was combined with the bandpass filter tuned to the signal frequency. The lowpass filter function also resembles a general loss in sensitivity to high-frequency amplitude modulations according to the data described in Kohlrausch *et al.* (2000). Table 1 shows the best fitting Q-values and  $-3$  dB bandwidths in Hz (in parenthesis) for each of the patterns in Fig. 3.2.

The Q-values range from 1 (for the 16-Hz signal and the 5.5-kHz carrier) to 1.8 (for the 16-Hz signal and the 1.4-kHz carrier). The average Q-value across all eight patterns is  $1.26(\pm 0.23)$ . Except for the 16-Hz signal at 1.4 kHz ( $Q = 1.8$ ), all estimates lie very close to the average value. Thus, the envelope-frequency selectivity can be characterized well by an estimated Q-value of about 1.25. Filter shape and Q-value do not depend systematically on either the signal-modulation frequency or the carrier frequency for the range of parameters tested here. The results are in qualitative agreement with the findings in Ewert and Dau (2000). They estimated the modulation-filter shape and bandwidth using the same experimental paradigm for broadband noise carriers and found an average



Q-value of about 1. However, their estimates showed somewhat more variability across the conditions. The present data are also consistent with the modulation masking data by Houtgast (1989) and Bacon and Grantham (1989) obtained with a different experimental paradigm than that used in the present study. Both studies used a noise carrier and fixed the masker (center) frequency while the signal frequency was varied. The present data are also in agreement with the results of Lorenzi et al. (2001a) using a different paradigm (see discussion).

### III. EXPERIMENT 2: THE ROLE OF ENVELOPE BEATS IN MODULATION MASKING

#### A. Rationale

This experiment investigates modulation masking obtained with different stimulus type combinations of the signal and the masker. While Experiment 1 examined tone-in-noise (TN) masking, the current experiment also investigates noise-in-tone (NT) as well as tone-in-tone (TT) masking. The question is whether effects similar to those found in the audio-frequency domain, as described in the introduction, can be observed in the envelope-frequency domain.

#### B. Method

##### 1. Subjects

Three subjects participated in the experiment. Two of the subjects (S1 and S3) also participated in Experiment 1. The third subject (S4) was the second author (JV). All subjects had clinically normal hearing and had experience in other psychoacoustic experiments. Their age ranged from 27 to 32 years. Subject S1 was paid for his participation on an hourly basis.

##### 2. Apparatus, stimuli and procedure

The stimuli were presented diotically via AKG K501 headphones. Subjects were seated in a sound-attenuating booth. The stimuli were generated digitally at a sampling rate of 48 kHz and converted to analog signals by a two-channel 24-bit DAC including reconstruction filtering (SEKD ADSP 2496). Signal generation and presentation during the experiments were computer controlled using the AFC (alternative forced choice) software package for MATLAB, developed at the University of Oldenburg. Modulation detection thresholds were obtained using a sinusoidal carrier of 2.8 kHz. The stimuli had a duration of 500 ms including 50-ms  $\cos^2$  ramps. signal and masker modulation were either a tone or a narrowband noise with a bandwidth of 22.3 Hz. Several signal-masker combinations were considered: (i) tone-in-noise masking (TN), where a sinusoidal signal modulation was masked by a noise; (ii) tone-in-tone masking (TT), where a sinusoidal signal was

masked by a sinusoidal masker; and (iii) noise-in-tone masking (NT), where a noise signal was masked by a tone. The signal modulation was always centered at 64 Hz. Thus, as in Experiment 1, the noise had a bandwidth of 1/2 octave in the on-frequency condition where the center frequency of the masker was equal to the signal frequency. The masker was centered at -2, -2/3, -1/3, 0, 1/3, 2/3 octaves relative to the signal. The masker rms level was set to -12 dB. In one additional experiment with a tonal signal and a tonal masker modulation (TTr), the masker level was roved by  $\pm 3$  dB. The roving of the modulation depth introduces a random power fluctuation to the pure-tone masker across the presentation intervals. This is more comparable to the narrowband-noise masker condition, while in contrast to the noise masker, the pure-tone masker still exhibits no inherent fluctuations. In contrast to Experiment 1, where a multiplicative approach was used, the sum of the masker and signal modulation was applied to the carrier. This results in the following equation for the stimuli:

$$s(t) = a \{ \sin(2\pi f_c t) [1 + m m_s(t) + b m_m(t)] \} , \quad (3.2)$$

where  $f_c$  is the carrier frequency,  $a$  is the amplitude of the stimulus,  $m$  indicates the signal-modulation depth, and  $m_s(t)$  represents the signal-modulation waveform.  $b$  is the masker modulation depth and  $m_m(t)$  the masker modulation waveform. Setting  $b$  to zero eliminates the masker modulation, as used to determine the unmasked reference threshold. The sinusoidal signal modulation in the TN, TT, and TTr condition as well as the sinusoidal masker modulation in the NT condition always started in sine phase. In the two tone-in-tone conditions (TT and TTr), the phase of the masker modulation was randomized. The overall presentation level was 65 dB SPL. The increase in level due to the presence of the signal and/or masker modulation was not compensated for<sup>1</sup>. As in Experiment 1, a three-interval, three-alternative forced-choice paradigm in combination with a one-up two-down tracking procedure was used to obtain threshold estimates.

## C. Results

The results were similar across subjects, so only the mean data are presented here. The upper panel of Fig. 3.3 shows results for the TN condition (circles) together with the data from Experiment 1 (replotted from Fig. 3.2), indicated as squares connected by the dashed lines. Masked thresholds are shown as a function of the masker shift (in octaves) relative to the signal (center) frequency (64 Hz). Values are given in dB rms of the modulator waveform in contrast to Experiment 1, where the modulation depth,  $m$ , in dB was given.

---

<sup>1</sup>By the stochastic nature of the narrowband noise signal and masker modulation, the individual noise samples in the different intervals may vary in their rms level, with a long-term average equal to -12 dB. This introduces slight level variations to the stimuli when applied to the fixed level pure-tone carrier. In order to avoid changes to the stimulus statistics, this increase in level due to the presence of the signal and/or masker modulation was not compensated for. It is, however, unlikely that the change in level caused by the signal modulation (< 0.25 dB) was used as the detection cue.

This was done since  $m$  is well defined only for a sinusoidal modulation<sup>2</sup>. In this experiment, a smaller stepsize (1/3 octave instead of 2/3 octaves) and a smaller spectral range of masker positions than in Experiment 1 was used (as indicated by the dashed box in the middle panel of Fig. 3.2). The solid horizontal line indicates the threshold in the absence of the masker modulation (unmasked threshold). The highest masked threshold can be observed for a masker frequency slightly lower than the signal frequency (-1/3 octave). Masking, defined as the difference between masked and unmasked threshold, amounts to 15.5 dB in this condition. Within the range from -2/3 to +2/3 octaves, masking is smallest (11.3 dB) for the +2/3 octave condition. For the masker presented two octaves below the signal tone, masking is further reduced to 3.2 dB.

The middle panel of Fig. 3.3 displays the results for the NT masking condition. While the unmasked threshold for the noise (horizontal line) is only 0.8 dB lower than that for the tone (solid horizontal line in the upper panel), the masked thresholds (circles) are considerably lower than in the TN condition shown in the upper panel. The amount of masking reaches a maximum of 9 dB for the on-frequency masker, i.e., about 6 dB less than in the TN condition. Thresholds show less variation with masker frequency than in the TN condition. For masker frequencies in the range from -2/3 to +2/3 octaves relative to the signal frequency, thresholds are only 3-5 dB higher than for the masker position at -2 octaves.

Finally, the lower panel of Fig. 3.3 shows the data for a sinusoidal signal modulation in the presence of a sinusoidal masker modulation. Data for two different masker conditions are shown: the circles indicate thresholds for a fixed masker level (TT) while the squares represent thresholds where the masker level was roved (TTr). The two masking patterns are very similar and do not differ by more than 2 dB from each other, except for the on-frequency condition where the difference is 3.8 dB. For all masker frequencies below and above the signal frequency, thresholds are roughly independent of the masker position. A substantial increase in threshold (of about 17 dB) is only observed for the on-frequency condition where the masker frequency equals the signal frequency. In this case no beating between masker and signal modulation can occur and the only cue for the presence of the signal modulation is a variation in the overall modulation depth at the signal frequency. Since signal and masker modulation are added in random phase, the sum will result in an increase or a decrease of the overall modulation depth, depending on the phase relation between the two. This effect leads to the elevated threshold in the on-frequency condition. The level roving in the TTr condition further increases the on-frequency threshold.

---

<sup>2</sup>Note that, for the sinusoidal signal modulation, the values in dB rms are 3 dB lower than the values expressed as modulation depth,  $m$ , in dB.

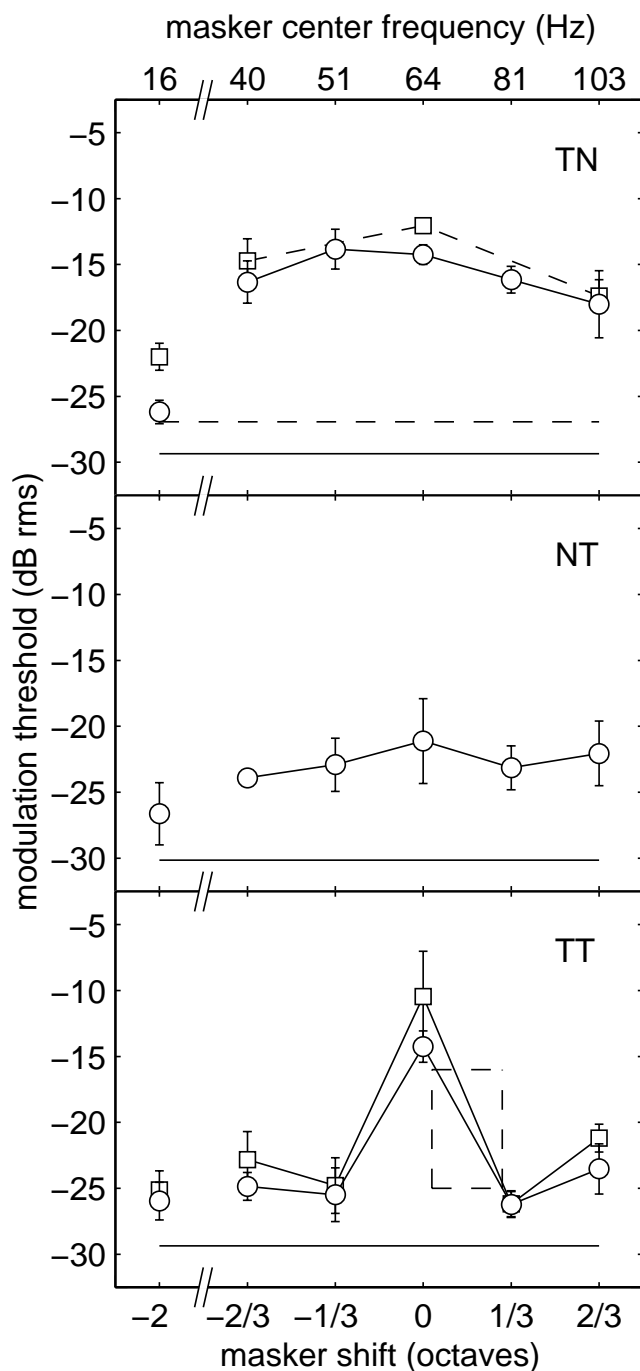


FIG. 3.3: Average masked-threshold patterns for a 64-Hz signal modulation as a function of the spectral position of the masker modulation. Panels show data for the tone-in-noise condition (TN, upper panel), noise-in-tone (NT, middle panel), and tone-in-tone (TT, lower panel). The lower panel shows also data for the TT condition with a 3-dB modulation-depth roving applied to the masker tone (indicated by the squares). Unmasked (reference) thresholds for the 64-Hz signal are represented by the horizontal line in all panels. The data points connected by the dashed lines and the horizontal dashed line in the upper panel are replotted from Fig. 3.2. The dashed box indicates the focus of the third experiment.

## IV. EXPERIMENT 3: INTERFERENCE OF ENVELOPE AND VEVELOPE PROCESSING IN TONE-IN-TONE MASKING

### A. Rationale

The previous experiment showed that thresholds in the tone-in-tone (TT) masking conditions are considerably lower than in the tone-in-noise (TN) masking conditions. In the TT conditions, the subjects probably use beats between the signal and the masker as an additional detection cue. Experiment 3 examines this hypothesis by adding a sinusoidal

modulation component with a period equal to the beat period between signal and masker in the TT condition. If the detection of beats is responsible for the threshold difference between the TT and TN condition, the additional component should influence the results.

## B. Method

Subjects, apparatus and procedure were the same as in Experiment 2. Thresholds for signal modulations of 32, 64, and 128 Hz were obtained in two masking conditions. The first one was the TT condition of Experiment 2, i.e., the masker was a pure-tone with an rms level of -12 dB. However, only spectral masker positions *above* the signal frequency were used (as indicated by the dashed box in the lower panel of Fig. 3.3 for the 64-Hz signal modulation). For a signal frequency of 128 Hz, the difference between the signal and the masker frequency was 2, 4, 8, 16, or 32 Hz, respectively. For signal frequencies of 32 and 64 Hz, the masker-signal separation was restricted to 2 to 8 and 2 to 16 Hz, respectively. In the second condition (TTm), the masker consisted of two components. The primary component was the same as used in the first condition. In addition, a second component was presented at the difference frequency between the signal and the masker. The level of this second component was roved within  $-\infty$  dB (absence) and -12 dB (uniform distribution on a linear scale) in order to prevent the subjects from using any reliable information based on the absence or presence of a temporal fluctuation at the masker-signal difference (beat) frequency. The equation for the stimuli in the two conditions was

$$s(t) = a \{ \sin(2\pi f_c t) [1 + m m_s(t) + b m_m(t) + r m_i(t)] \}, \quad (3.3)$$

where  $r$  is the modulation depth of the random-phase interfering tone  $m_i(t)$  at the difference frequency. The other abbreviations are the same as used in Equation 3.2. For the first condition (TT),  $r$  was equal to zero. For the second condition (TTm),  $r$  was chosen randomly in each realization.

## C. Results

Figure 3.4 shows average masked thresholds as a function of the spectral separation between (primary) masker and signal modulation. The three panels display results for signal frequencies of 128 Hz (top), 64 Hz (mid) and 32 Hz (bottom). The circles represent the thresholds for the TT condition. The squares indicate the thresholds obtained in the presence of the additional interferer at the difference frequency (TTm). The position of the masker was shifted in certain steps above the signal frequency, e.g., for the 128-Hz signal frequency (top panel), the resulting masker frequencies were 130, 132, 136, 144, and 160 Hz, respectively. The middle panel additionally shows the unmasked threshold for the 64-Hz signal as a horizontal line. The triangle indicates the threshold when only the interferer tone at 16 Hz was presented. As shown in Kohlrausch *et al.* (2000), very similar unmasked thresholds can be expected for the other two signal frequencies.

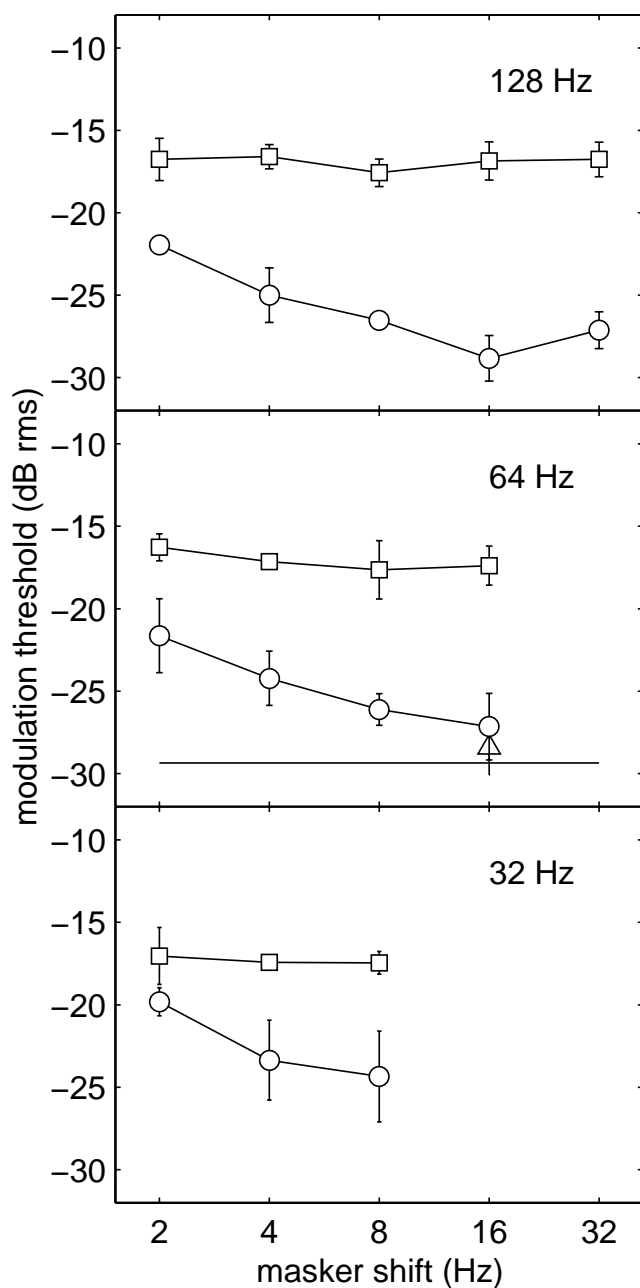


FIG. 3.4: Average masked-threshold patterns (circles) for signal-modulation frequencies of 128 Hz (upper panel), 64 Hz (middle panel), and 32 Hz (lower panel). The masker was always higher in frequency. The frequency difference between masker and signal is indicated at the bottom and ranged to a quarter of the signal frequency in all conditions. The squares represent masked thresholds in presence of an additional masker modulation at the beat frequency of the signal and primary masker. The horizontal line in the middle panel indicates the unmasked threshold for the 64-Hz signal modulation.

The threshold patterns are similar for the three signal frequencies. In the TT conditions (circles), masked threshold decreases with increasing masker shift, except for the highest masker shift (32 Hz) for the 128-Hz signal frequency, where threshold is slightly increased (by about 2 dB) compared to the 16-Hz shift. For a masker shift of 16 Hz, threshold almost reaches the unmasked threshold for 64 and 128 Hz signal frequency. For all three signal frequencies the shape of the threshold pattern strongly changes when the interfering modulation at the masker-signal difference frequency (beat rate) is presented in addition to the signal and the (primary) masker modulation (indicated by the squares). A threshold of about  $-17$  dB is obtained for all masker positions and all signal frequencies. This threshold is only 2.8 dB below the masked threshold obtained in the TN condition for the on-frequency masker (see Fig. 3.3). The average threshold difference between the two

curves amounts to about 9 dB for 8-Hz masker shift and signal frequencies of 64 and 128 Hz and to about 7 dB for 32 Hz signal frequency. The maximum difference of 12 dB between the thresholds in the two conditions is observed for a masker shift of 16 Hz. The threshold increase cannot be explained by the masking effect caused by the interferer alone since the threshold for the 64-Hz signal frequency in the presence of a 16-Hz tonal interferer alone, indicated by the triangle, is only slightly higher than the unmasked threshold.

## V. DISCUSSION

The primary purpose of the present study was to examine the extent to which the detection of envelope beats, or envelope fluctuations, influences modulation detection. The question was whether the detection of envelope fluctuations plays a similar role in modulation masking as does the detection of envelope fluctuations in spectral masking. While the envelope of a stimulus is extracted by the halfwave-rectifying properties of the inner hair cells, it is unclear which process may cause a “demodulation” of the envelope to the envelope domain. In the following, first the experimental results of the present study are discussed in the context of recent literature data. In the second part, a general modeling structure for the processing of the envelope and the envelope in the auditory system is proposed.

### A. Experimental results

Experiment 1 examined envelope-frequency selectivity using pure-tone carriers and served as a reference for the subsequent experiments. Masked threshold patterns for three different signal-modulation frequencies (4, 16, and 64 Hz) were obtained for three carrier frequencies. The similarity of the Q-values of the filters fitted to the data, suggest that modulation masking can be accounted for very well by assuming an *invariant* array of modulation filters at the output of each critical band. Alternatively, the results may, in principle, also support the hypothesis that envelope information is first integrated over a wide range of audio-frequencies and is then subjected to a *common* modulation filterbank. Neither of the two possible processing schemes can be excluded from the experimental findings. However, experiments on modulation detection interference (MDI) showed that results can not easily be modeled in terms of “hard-wired” modulation cross-talk across peripheral frequency channels (Hall and Grose, 1991; Oxenham and Dau, 2001). The analysis of across-frequency information seems to take place only after some degree of perceptual grouping has occurred, suggesting that information is integrated at some more central stage of processing. It therefore seems plausible to assume that envelope information is first processed separately (at different carrier frequencies) in terms of a modulation filterbank analysis and is later combined across frequency in a way depending on the specific acoustical context of the stimuli. The assumption of an invariant modulation filterbank mechanism whose parameters do not change with carrier frequency was already made in the first model implementation described in Dau *et al.* (1997a,b).

Experiment 2 investigated modulation masking patterns obtained for one of the signal-modulation frequencies (64 Hz) and one of the carrier frequencies (2.8 kHz) used in Experiment 1 whereby only spectral masker positions *close* to the signal frequency were considered. In these conditions, the low-frequency beats between the signal and the masker provided a strong detection cue depending on the stimulus type of the signal and the masker. The masking patterns obtained for TN, NT, and TT masking conditions showed the same general characteristics as masking patterns obtained in the audio-frequency domain (e.g., Moore *et al.*, 1998): A narrowband noise is more effective in masking a tonal signal (TN) than vice versa (NT). Also, if a tonal signal is used, a narrowband noise (TN) is a more effective masker than a tone (TT). The results are especially striking in the TT condition, where only little masking is observed, except for the on-frequency condition. However, the on-frequency condition of the TT masking pattern is comparable to a modulation-depth discrimination experiment and thus does not provide any information on modulation masking. In the framework of a model that is restricted to only the mean envelope power passing the modulation filter tuned to the signal frequency (EPSM, Ewert and Dau, 2000), the same pattern of results would be expected for all masking conditions. Such a model would not include effects of temporal envelope fluctuations originating from beats between the components of the signal and masker modulation. A masking pattern as expected from the EPSM is empirically observed only for the TN condition. This suggests that, in contrast to the TN condition, the envelope provides additional detection cues in the TT and the NT condition, similar to the envelope in the audio-frequency domain. However, it is important to realize that in the audio-frequency domain, envelope fluctuations (reflecting beats between the spectral components) and resolved spectral components have a completely different perceptual quality while in the modulation domain, envelope fluctuations (reflecting beats between envelope components) and “resolved” envelope components are of the *same* perceptual quality. In the latter case both evoke a temporal impression. Thus, there are two independent “dimensions” (perceptually and neurally) represented in the periphery, the frequency and envelope frequency axes, while there is no further independent dimension available for the coding of envelope fluctuations.

This was further supported by the results from Experiment 3 where an additional modulation component at the difference frequency between the (primary) masker and the signal was imposed in a tone-in-tone masking configuration. The substantial threshold elevation observed in the presence of such an “interferer” clearly demonstrated that beats must have been the prominent detection cue in the TT masking condition (without the additional component). As demonstrated for the 64-Hz signal and a 16-Hz interferer, the masking effect caused by the interferer alone cannot account for the threshold increase. This should hold especially for the larger spectral separations between the signal and the interferer. Since the Q-value of the modulation filters is independent of filter center frequency, the same is expected for signal frequencies of 32 and 128-Hz. The data clearly showed that the detection of a modulation strongly interferes with the beating modulators fluctuating at the same rate. This experiment differs from the experimental paradigm in



Moore *et al.* (1999) since a low-frequency component was introduced in order to mask the beat cue produced by two higher frequency components while Moore *et al.* study investigated masking of a low-frequency component by introducing two higher frequency beating components. Nevertheless, the underlying mechanisms are probably the same in the two studies. In addition, the findings in Experiment 3 suggest that beats become less prominent at low rates (2 Hz) than at higher rates (8 to 16 Hz). This is probably caused by the reduced number of beat cycles available during the stimulus presentation of 500 ms.

Lorenzi *et al.* (2001b) suggested that the detection of envelope fluctuations might *not entirely* be based on the detection of the physical envelope component(s) introduced by some nonlinear process. They measured “second-order” modulation detection thresholds with pure-tone and 2-Hz wide narrowband noise carriers. In contrast to the pure-tone carrier, the narrowband noise carrier exhibits low-frequency intrinsic envelope fluctuations (Lawson and Uhlenbeck, 1950; Dau *et al.*, 1997a,b). If envelope detection was entirely based on the detection of energy in the envelope domain, thresholds for low-frequency envelopes should be considerably higher in case of a narrowband noise carrier than in case of a pure-tone carrier. Even though some masking was observed in that study the amount of masking was only relatively weak compared to that observed for first-order modulations. The authors suggested that possibly a second “unknown” process might be involved in the detection of the envelope, suggesting that the envelope may be represented “independently” at some intermediate stage of processing.

Overall, the findings from the present study clearly suggest that (i) the envelope is extracted by the auditory system at some stage of processing and that (ii) envelope and envelope fluctuations evoke the same (temporal) perception. This suggests a common encoding scheme for the processing of the envelope and the envelope of the stimulus.

## B. Possible model structures

Figure 3.5 shows a general processing scheme that can in principle account for the empirical results from the present study. In parallel to the envelope path, a second path extracts the envelope of the stimulus. The mathematical definition of the envelope is given in Appendix B. After some amount of attenuation (which is a free parameter in this scheme) the envelope is added to the envelope path. The resulting activity is then subjected to a modulation filterbank. Thus, it is assumed that the extraction of the envelope and the combination of envelope and envelope occur at a processing stage prior to the bandpass filtering process. This processing scheme represents a very simple and straight-forward way to realize *functionally* an internal representation that contains envelope and envelope in the same “dimension”. Such a scheme is also compatible with recent data from Lorenzi *et al.* (2001a) on second-order modulation detection. Lorenzi *et al.* (2001a) concluded from their data that, if modulation filters do exist, they have to be broadly tuned (Q-value less than 2) in order to avoid the detection of spectrally resolved envelope frequencies in different modulation filters. The current filters have a Q-value of about 1.

Physiological mechanisms that may generate neural activity at frequencies correspond-

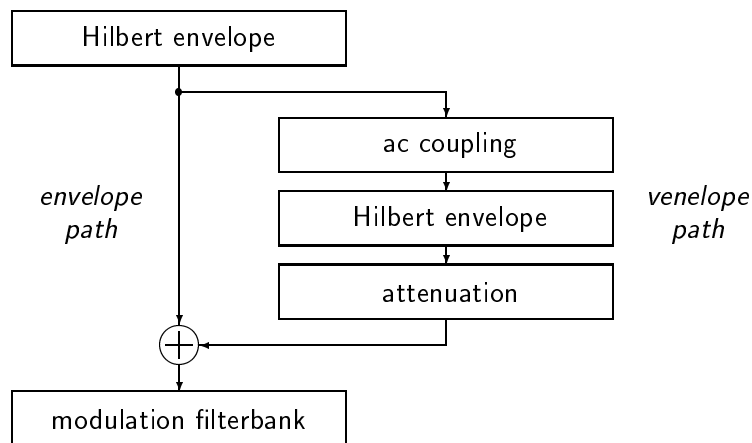


FIG. 3.5: General processing scheme to account for the empirical results from the present study. In parallel to the envelope path, a second path extracts the venelope (second-order envelope) of the stimulus. After attenuation (free parameter in this scheme) the venelope is added to the envelope. The resulting “envelope” is then subjected to a modulation filterbank.

ing to the venelope frequency have been discussed recently in the literature. For example, Shofner *et al.* (1996) measured responses of neurons in the chinchilla cochlear nucleus to carriers at the characteristic frequency that were amplitude modulated by two-component modulators. The study revealed evidence for a distortion component at the beat rate of the modulators. In the following, some possible mechanisms for the generation of such a component are compared to the venelope concept.

The top panels of Fig. 3.6 show the temporal envelope of a two-tone modulator with components of 64 and 80 Hz (left; compare to Fig. 3.1) and the corresponding envelope spectrum (right). There is no energy at the beat rate of 16 Hz in the stimulus envelope. The remaining panels show corresponding stimuli and spectra obtained with different transformations all of which introduce energy at the beat rate. In addition to the envelope waveforms, the left panels also show the waveform of the component at the difference frequency. Since the energy of this component is rather small compared to that of the primary components, the waveform was multiplied by a factor of ten for illustration. The panels in the second row show the effect of a *compressive* nonlinearity as suggested in some studies (Shofner *et al.*, 1996; Sheft and Yost, 1997; Moore *et al.*, 1999). In this case, a power-law function with an exponent of 0.4 was assumed. In the third row, the stimulus envelope was clipped at a certain threshold level, as suggested by Shofner *et al.* (1996). The fourth row displays the waveform and the spectrum at the output of the venelope model as proposed in the present study, assuming an arbitrary attenuation factor of 0.2. Finally, the bottom panel shows the envelope calculated at the output of a critical band filter tuned “off-frequency” to the lowest sideband originating from the modulation. All processing schemes clearly generate a physical envelope component at the beat frequency. The size of this distortion component depends on the transformation. For all transformations except the off-frequency filter scheme (bottom panel) the size of

the component at the difference frequency scales with the parameters given above, i.e, the exponent, the threshold value or the attenuation factor. The off-frequency filter scheme produces a much weaker component compared to the nonlinear transformations. Probably more importantly, the phase of the introduced component relative to the phase of the beat differs for the different transformations. A compressive nonlinearity results in a 180-degree phase shift relative to the phase of the envelope. In contrast, all other transformations produce a component in phase with the envelope, as would do an expansive nonlinearity (not shown) in the processing path. The off-frequency calculation nicely shows that it is not necessarily a nonlinear mechanism that is needed to demodulate the beat component to the envelope domain. Since the data of the present study do not depend on the relative phase between the different modulator components they do not allow one to distinguish between the different model realizations. Critical experiments that are focused on these phase effects are described in a companion paper (Verhey *et al.*, 2002). Their results suggest that a compressive nonlinearity, such as that introduced by the input-output function of the basilar membrane (Sellick *et al.*, 1982; Ruggero *et al.*, 1997; Moore and Oxenham, 1998), does not account for their modulation beating data. This agrees with results from a recent study by Tandetnik *et al.* (2001), where it was found that second-order modulation detection thresholds were very similar in sensorineural hearing-impaired listeners and in normal-hearing listeners. Cochlear damage results in a reduction or loss of the compressive nonlinearity in the input-output function of the basilar membrane (see Moore, 1995, for a review). This suggests that the detection of envelope fluctuations is not related to nonlinear processing on the basilar membrane.

While the above mentioned modeling schemes assumed the introduction of spectral energy at the envelope beat rate *prior* to the modulation bandpass filtering process, it is in principle also possible that low-frequency envelope beat cues appear in the temporal pattern at the output of modulation filters tuned to the (higher) signal and masker modulation rates. Theoretically, such envelope beat cues should then not be affected by the presentation of an interfering modulator at the beat rate if the signal/masker modulation rates and the beat rates are sufficiently separated from each other. Thus, in order to predict a masking effect as observed in the data, some integration/interference mechanism across envelope frequencies would have to be assumed *after* the modulation filtering process. Such processing certainly reflects a possible alternative modeling scheme within a more complex model that preserves and analyzes the temporal structure of the output of modulation filters (unlike the EPSM). The envelope/envelope processing scheme proposed in the present study may be considered as the most straight-forward and functional model which is consistent with the experimental data from the present study. Although the envelope and envelope path are combined in order to account for the empirical data, one main feature of the envelope extraction scheme is the parallel representation of the envelope and envelope at some stage of the model, in contrast to pure “inline” schemes such as the compression model by Moore *et al.* (1999). Future investigations will show how useful this approach is for the description of the internal representation(s) of the stimuli

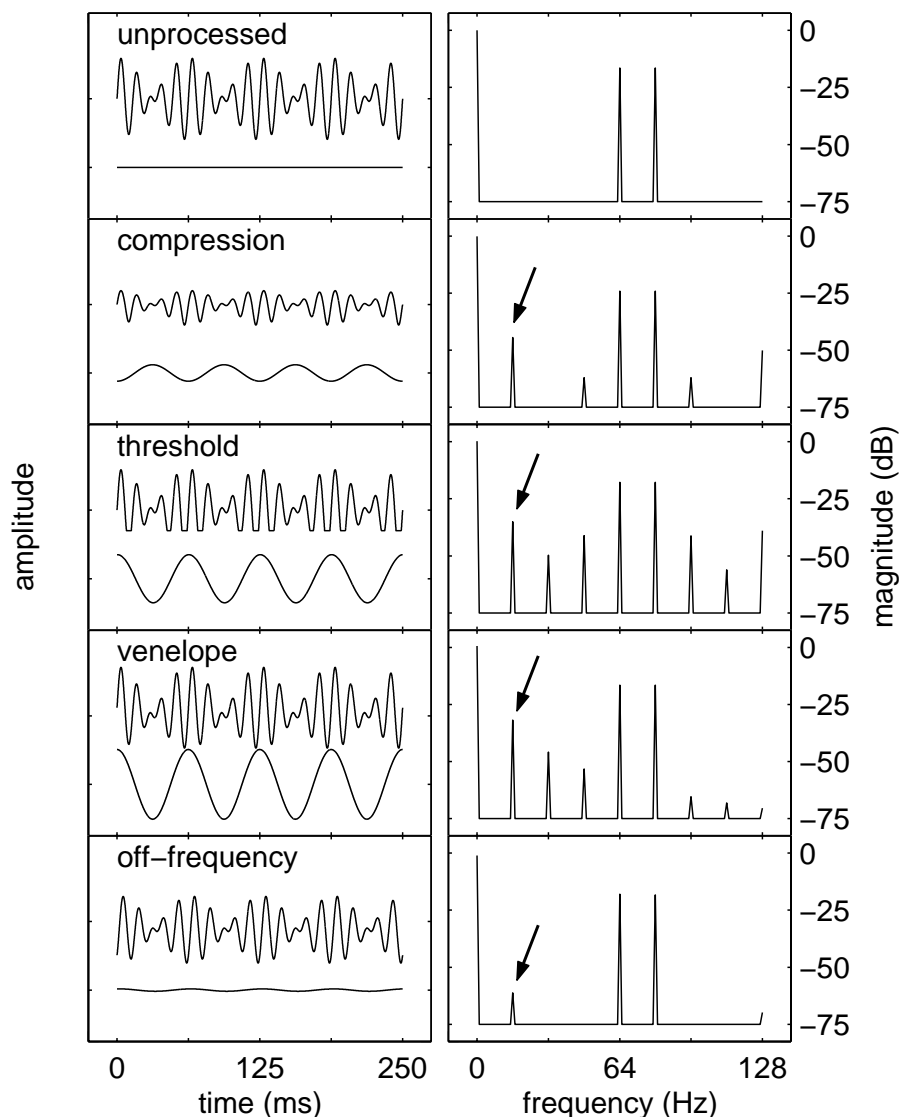


FIG. 3.6: Each panel of the left hand column shows the envelope waveform (upper graph) resulting from the sum of two sinusoidal amplitude modulations at 64 and 80 Hz. The lower graph in each of the panels is the 16 Hz component of the envelope multiplied by a factor of 10. The right hand column shows the corresponding envelope spectra. Different transformations were applied to the envelope in each row, as indicated in the left hand panels.

in simple and complex stimulus configurations.

## VI. SUMMARY AND CONCLUSIONS

The results of this study provide further evidence that linear spectral analysis of the (first-order) envelope is not sufficient to describe all temporal properties of non-stationary stimuli that are relevant for our perception. As the “synthetic” stimuli used in psychoacoustical experiments, most natural sounds are not limited to a single component in the envelope spectrum. Indeed, typical everyday-life sounds like speech (Plomp, 1988; Drullman et al., 1994; Greenberg and Arai, 1998) and noise-like sounds (e.g., Lawson and

Uhlenbeck, 1950) exhibit a number of components in the envelope spectrum. Since the auditory system most likely processes the envelope nonlinearly, distortion components are introduced to the internal representation of the stimuli. These components were demonstrated to be perceptually highly relevant in specific modulation masking experiments. The main results of the study are as follows:

- Peaked masked-threshold patterns (MTPs) were obtained for signal-modulation frequencies in the range from 4 to 64 Hz and a half-octave-wide noise masker, using pure-tone carriers (Experiment 1). The shape of the MTPs did not depend on the carrier frequency. The estimated Q-value of the filters was about 1.2. This is in good agreement with the Q-value of 1 found in a previous study with broadband noise carriers (Ewert and Dau, 2000) using the same envelope power spectrum model and the same fitting procedure.
- Masked threshold patterns showed large threshold variations for different combinations of sinusoidal and narrowband noise masker and signal modulators (Experiment 2). Thresholds in noise-in-tone (NT) masking conditions were always lower than in tone-in-noise (TN) masking, similar to corresponding experiments in the audio-frequency domain. In the NT conditions, subjects use beats between the noise signal and the tone masker as a detection cue, while in the TN conditions, the beats are masked by the inherent fluctuations of the noise masker.
- In tone-in-tone (TT) masking conditions, thresholds were close to the unmasked threshold for spectral separations  $\geq 4$  Hz since the beats produced by the signal and the masker provided a strong detection cue. By adding an additional sinusoidal modulation at the difference frequency, the salience of the beats was strongly reduced and thresholds were raised up to 12 dB.
- The venelope defined as the Hilbert envelope of the ac-coupled (first-order) Hilbert envelope provides an analytical tool for describing the interactions between the different envelope components for arbitrary multi-component waveforms. A model structure that combines the envelope and the venelope of a stimulus was proposed to qualitatively account for the data presented here. The venelope concept is only one possible “mechanism” that produces distortion. A set of physiologically plausible nonlinearities were discussed. A critical test of these is undertaken in an accompanying study (Verhey *et al.*, 2002).

## ACKNOWLEDGMENTS

We thank Neil Viemeister and two anonymous reviewers for very helpful suggestions and Andrew Oxenham for comments on an earlier version of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG).



## CHAPTER 4

# MODULATION MASKING PRODUCED BY COMPLEX TONE MASKERS<sup>a</sup>

### ABSTRACT

Thresholds were measured for detecting sinusoidal amplitude modulation in the presence of a complex-tone masker modulation. Both modulations were applied to a 5-kHz sinusoidal carrier. Two different masker modulators were used: i) a pair of components beating at the difference frequency and ii) a three-tone complex producing a sinusoidal amplitude modulation of the modulation depth at the difference frequency between adjacent components. Thus, both maskers show a periodicity in the envelope waveform that is not contained in the envelope spectrum itself but can be observed when the envelope of the envelope, referred to as the “venvelope” [Ewert et al., *J. Acoust. Soc. Am.*, in press], is calculated. For a signal frequency equal to the masker venelope periodicity, modulation depth at threshold was measured as a function of the signal phase relative to the phase of the masker venelope component at the signal frequency. A signal frequency of 30 Hz was used. Thresholds were lower for the in-phase condition, where the maxima in the signal waveform coincide with the maxima in the venelope waveform, than for the anti-phase condition. The maximum threshold difference was 15 dB. The same phase effect was obtained in normal-hearing and sensorineural hearing-impaired listeners. The results are in contrast to recent data [B.C.J. Moore et al., *J. Acoust. Soc. Am.* **106**, 908-918 (1999)], where lowest thresholds were found for the anti-phase condition. The present data are in line with the idea that a nonlinearity prior to a modulation filterbank extracts the venelope of the masker modulator. A compressive nonlinearity, however, as associated with the processing on the basilar membrane, cannot account for the empirical findings. It

---

<sup>a</sup>This chapter is a modified version of J. L. Verhey, S. D. Ewert, and T. Dau, (2002). “Modulation masking produced by complex tone modulators”, submitted to *J. Acoust. Soc. Am.* .

predicts a phase effect opposite to that observed in the data and is assumed to be reduced or absent in hearing-impaired listeners.

## I. INTRODUCTION

The concept of critical bands, first introduced by Fletcher (1940), has been successfully used to explain detection performance in the audio-frequency domain. It assumes that the auditory system analyzes the signal by a bank of overlapping bandpass filters and the signal-to-noise ratio in the filter at the signal frequency determines threshold. More recent studies have shown that the auditory system may use a similar mechanism for envelope fluctuations. Psychoacoustical modulation detection experiments using narrow-band noise carriers (Fleischer, 1982, 1983; Dau *et al.*, 1997a; Lorenzi *et al.*, 2001a), modulation masking data with narrow-band noise masker modulators (Houtgast, 1989; Bacon and Grantham, 1989; Ewert and Dau, 2000; Ewert *et al.*, 2002) and physiological data (Langner and Schreiner, 1888; Palmer, 1995; Schulze and Langner, 1997, 1999) suggest that the auditory system may also perform a frequency-selective analysis in the envelope-frequency domain.

Although several researchers have proposed such a mechanism (e.g., Martens, 1982; Houtgast, 1989; Bacon and Grantham, 1989; Dau *et al.*, 1997a), there are only a few studies where quantitative predictions with a model including a modulation filterbank were presented. Dau *et al.* (1997a) expanded their original signal processing model (Dau *et al.*, 1996) by introducing a modulation filterbank at the output of each peripheral filter. It was shown that this model can account for a large variety of detection and masking data in both the audio-frequency and the envelope-frequency domain (Dau *et al.* 1997a, b; Verhey *et al.*, 1999; Derleth and Dau, 2000; Verhey, 2002). It was also shown that a more simplified model, only regarding the integrated envelope power at the output of a modulation filter (Dau *et al.*, 1999, Ewert and Dau, 2000, Ewert *et al.* 2002), can account for modulation masking data obtained for noise and pure-tone carriers.

Recently, Moore *et al.* (1999) found that the detection threshold for a 5-Hz signal modulation was affected by the presence of a spectrally remote pair of masker modulators. Masker modulators of 40 and 45 Hz (and higher) were used, beating at the rate of the signal modulation. They found that the signal modulation depth at threshold depends on the phase of the signal relative to the beat cycle. At first sight, this seems to be in contrast to the concept of modulation filters, suggesting that envelope components can only interact if they excite a common modulation filter. However, Moore *et al.* (1999) argued that the compressive nonlinearity associated with the processing on the basilar membrane (e.g., Ruggero, 1992) introduces a physical distortion component at the beat frequency in the internal representation of the envelope which interacts with the signal modulation. They showed that their data can be predicted within a model assuming a compressive nonlinearity prior to a modulation filterbank.

To characterize variations in the modulation depth of complex modulators in more gen-



eral terms, Ewert *et al.* (2002) suggested the “venelope”, which is defined as the Hilbert envelope of the ac-coupled Hilbert envelope of a stimulus. They showed that the venelope can be used as a detection cue in certain modulation masking experiments comparable to the “asymmetry of masking” effect in the audio-frequency domain (Hellman, 1972; Hall, 1997; Derleth and Dau, 2000; Verhey, 2002). Thresholds were markedly lower when a narrow-band noise was masked by a pure-tone masker than in the reversed condition. By analogy with the interpretation of the asymmetry of masking effect in the audio-frequency domain, where amplitude modulations provide important detection cues (Derleth and Dau, 2000; Verhey, 2002), they concluded that the venelope fluctuations provide a strong detection cue in corresponding experiments in the envelope domain. In addition, Lorenzi *et al.* (2001a,b) measured the detectability of a sinusoidal variation applied to the modulation depth of a “carrier” amplitude modulation. It was shown that in some conditions, the detection thresholds for the resulting venelope fluctuations were as low as those for envelope fluctuations.

Taken together, the above studies demonstrated the perceptual salience of the venelope of the stimuli. The data are in line with the idea that some nonlinearity extracts the venelope and combines it with the internal representation of the envelope. Compression, as proposed by Moore *et al.* (1999), is only one possible nonlinearity that could be used for extracting the venelope. For example, Shofner *et al.* (1996) proposed three different physiological mechanisms to account for their findings that units in the ventral cochlea nucleus showed a response to the beating of a two-tone complex modulation: compression, saturation and threshold. Ewert *et al.* (2002) proposed a black-box approach, assuming that the auditory system may extract the venelope of the stimulus in some way without specifying how this might be realized neurally. None of the previous studies investigated explicitly which of the proposed mechanisms accounts best for the psychophysical data.

The main goal of the present study is to investigate the underlying process for the extraction of the venelope in the auditory system. In order to further specify the nonlinear mechanism that may be responsible for the extraction of the venelope, venelope-phase sensitive experiments like those described in the Moore *et al.* (1999) study are necessary. Moore *et al.* (1999) showed the effect of the masker venelope phase on the threshold of a signal modulation only for a modulation frequency of 5 Hz. Only one modulation depth for each masker component was tested and only masker components that are harmonics of the signal-modulation frequency were used. The first experiment of the present study examines the role of signal phase at a signal-frequency of 30 Hz. The experimental paradigm was comparable to the one presented in Moore *et al.* (1999). In addition, the spectral distance between signal and masker and the harmonic relation between the signal frequency and the masker components were varied. The signal modulation threshold was measured as a function of the relative phase between masker venelope and signal modulation for various spectral distances between the masker components and the signal modulation. The signal modulation always matched the beat frequency. The second experiment compares the effect of the relative phase between masker venelope and signal modulation in normal-

hearing listeners with those obtained in sensorineural hearing-impaired listeners. Cochlear damage is assumed to result in a reduction or loss of the compressive nonlinearity in the input-output function of the basilar membrane (see Moore, 1995 for a review). The experiment thus provides a critical test for the role of the compressive nonlinearity on the basilar membrane in nonlinear envelope processing. In this experiment, a sinusoidal envelope applied to a high-frequency masker was used, resulting in a three-tone complex masker. Compared to the two-tone masker used in the first experiment, such a masker has a well-defined envelope spectrum, consisting of a single component at the imposed envelope frequency. The effect of the phase was measured for various envelope levels (modulation depths of the complex tone masker) in order to determine the lower limit for the interaction between envelope and envelope.

In the second part of the study, threshold predictions of a model including envelope extraction, some nonlinearity, and a modulation filterbank were compared to the empirical data. The models differ in the nonlinearity assumed for the extraction of the envelope prior to the modulation filterbank.

## II. METHODS

### A. Subjects

Three normal-hearing subjects ranging in age between 25 and 31 years participated in both experiments. One of them was author JV. All subjects had normal audiograms (i.e., absolute threshold in quiet  $\leq 15$  dB HL) and no previous history of hearing disorders. All subjects had experience in psychoacoustic experiments. Two subjects were members of the “Arbeitsgruppe Medizinische Physik” at the Universität Oldenburg. One subject was a paid volunteer. Five sensorineural hearing-impaired subjects ranging in age between 42 and 68 years participated in the second experiment. Three of them participated in both parts of the experiment. One subject that participated in the first part was not available for the second part and was replaced. The subjects had a bilateral similar flat hearing loss in the region of the carrier frequency, one subject had a mildly downward sloping hearing loss. Their audiometric threshold at the carrier frequency ranged from 40 to 60 dB SPL. The threshold difference between bone- and air-conduction was smaller than 10 dB. Audiometric data on the hearing-impaired subjects is given in Table 4.1. All hearing-impaired subjects had prior experience in psychoacoustic experiments and were paid for their participation in the experiment.

### B. Apparatus and stimuli

The subjects were seated in a double-walled sound-attenuating booth. The stimuli were generated digitally at a sampling rate of 32 kHz. Stimulus generation and presentation were controlled with the software package SI developed at the Universität Göttingen, using a Silicon Graphics workstation (INDY), which also sampled the listener’s responses

TABLE 4.1: Audiometric data on the five sensorineural hearing-impaired listeners at the 3 kHz.

Subject	Age	Ear	Pure-tone hearing loss (dB HL)			Classification of loss
			Air conduction	Discomfort level	Carrier level	
MG	42	R	40	105	80	Flat
WF	66	R	60	110	85	Flat
WW	49	L	45	120	80	Flat
MW	58	L	50	105	80	Mildly downward sloping
LF	68	R	45	115	85	Flat

and controlled the procedure. The stimuli were converted to analog signals (16-bit DAC), and then preamplified and lowpass filtered at 16 kHz using a computer-controlled audiometric amplifier. Stimulus presentation was diotic via Sennheiser HD25 headphones. For the hearing-impaired subjects, the stimuli were generated digitally at a sampling rate of 48 kHz and converted to analog signals by a two-channel 16-bit DAC including reconstruction filtering (SEKD ADSP 2496). Signal generation and presentation during the experiments were computer controlled using the AFC package for MATLAB, developed at the Universität Oldenburg. The stimuli were presented monaurally via Sennheiser HD580 headphones.

A sinusoidal carrier at 5 kHz was used in both experiments for the normal-hearing subjects. The carrier was 600 ms in duration including 20-ms  $\cos^2$  ramps. The carrier level was fixed at 70 dB SPL for the normal-hearing subjects. For hearing-impaired subjects, the carrier frequency was 3 kHz and the carrier level was either 80 or 85 dB SPL (see Table 4.1). In all cases, the stimuli were well audible. The signal and masker modulations were applied during the whole period of the carrier. The expression describing the masker-alone waveform  $W_m(t)$  is

$$W_m(t) = C(t)(1 + M_m(t)), \quad (4.1)$$

and the expression describing the masker+signal waveform  $W_{m+s}(t)$  is

$$W_{m+s}(t) = C(t)(1 + M_s(t) + M_m(t)), \quad (4.2)$$

where  $C(t)$  is the sinusoidal carrier,  $M_s(t)$  is the signal modulator, and  $M_m(t)$  is the masker modulator. The sinusoidal signal modulation is described by the following expression:

$$M_s(t) = m_s \sin(2\pi f_s t + \phi_s), \quad (4.3)$$

where  $m_s$  is the signal modulation depth,  $f_s$  is the signal modulation frequency and  $\phi_s$  is the starting phase. Two different masker modulators were used.

i) In the first experiment, a pair of sinusoidal modulators was used. The expression describing the two-tone masker is:

$$M_m(t) = m_m \sin(2\pi f_{m_1}(t + t_0)) + m_m \sin(2\pi f_{m_2}(t + t_0)), \quad (4.4)$$

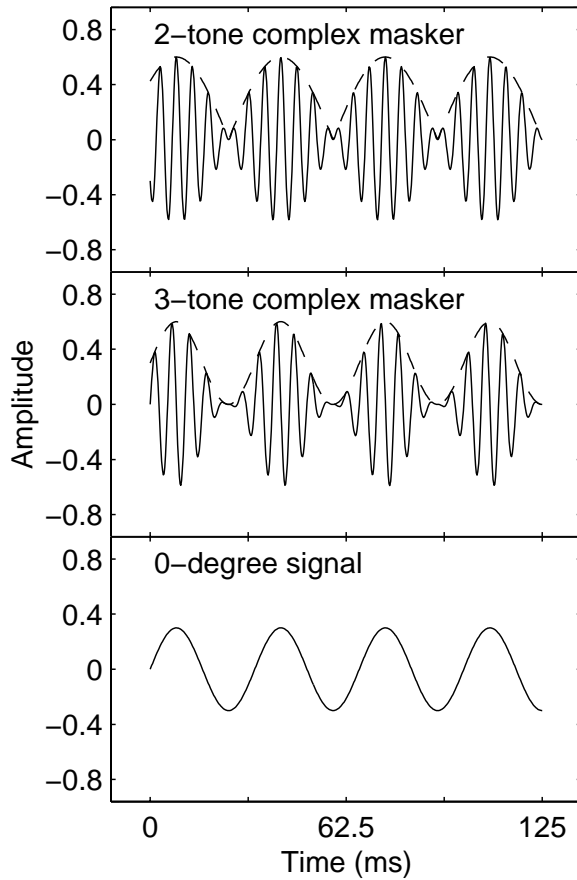


FIG. 4.1: The upper two panels show the temporal waveform of the 2-tone-complex masker (top panel) and the three-tone-complex masker (middle panel). In addition, the dashed line indicates the envelope for the two masker waveforms. The lower panel shows signal modulation in  $0^\circ$  starting phase. In this “in-phase” condition, maxima in the signal waveform coincide with those in the masker envelope.

where  $a_m$  is the modulation depth of each masker component,  $f_{m_1}$  and  $f_{m_2}$  are the frequencies of the two component and  $t_0$  is a constant time shift. The spectral distance between the two components is set to be equal to the frequency  $f_s$  of the signal modulation, i.e.,  $f_s = f_{m_2} - f_{m_1}$  ( $f_{m_2} > f_{m_1}$ ). The modulation depth  $m_m$  of each masker component was set to 0.3. This corresponds to a modulation depth of about -10 dB in terms of  $20 \log m$ . The time shift  $t_0$  was set to  $0.75/f_s$ . Except for the time shift, the equation for this masker is the same as used in Moore *et al.* (1999)<sup>1</sup>. As an example, the waveform of a two-tone complex masker modulator with  $f_{m_1} = 180$  Hz and  $f_{m_2} = 210$  Hz is shown in the upper panel of Fig. 4.1. The pair of modulation components produces beats at the difference frequency of 30 Hz. The envelope waveform of the two-tone-complex masker (dashed line) is a rectified sinusoid with a frequency equal to half the beat frequency. Thus, the spectrum of the envelope has a fundamental frequency component at 30 Hz but also contains higher order harmonics which are integer multiples of the fundamental frequency. Note that 30 Hz is not contained in the modulation spectrum itself, which only contains  $f_{m_1}$  and  $f_{m_2}$ . The time shift  $t_0$  was introduced in order to obtain a masker modulation with a envelope which has a maximum at the same time as the signal modulation with a starting phase of  $0^\circ$  (lower panel of Fig. 4.1). Accordingly, the maximum of the envelope

<sup>1</sup>Moore *et al.* (1999) defined the signal phase relative to the phase of the beating component. Apparently, this implies that in their study the signal cannot start in sine phase for a relative phase of  $0^\circ$ . In contrast, in the present study, the relative phase coincides with the starting phase of the sinusoidal signal.

corresponds with the minimum of the signal with a starting phase of  $180^\circ$  (not shown).

ii) In the second experiment, the masker modulation consisted of a modulation component at  $f_m$  that was sinusoidally modulated in its modulation depth. The equation describing the masker is:

$$M_m(t) = m_m \sin(2\pi f_m t)(1 + \sin(2\pi f_s t)), \quad (4.5)$$

where  $m_m$  is the mean modulation depth and  $f_s$  is the signal modulation frequency. Thus, the masker modulation was a tone with a frequency  $f_m$  which was 100% sinusoidally amplitude modulated at a rate that was equal to the signal frequency. In contrast to the 2-tone complex, the envelope of the masker contains only the signal frequency. As an example, the masker modulator waveform for  $f_m = 180$  Hz and  $f_s = 30$  Hz is shown in the middle panel of Fig. 4.1. The envelope is a sinusoid with a frequency  $f_m$  of 30 Hz (indicated by the dashed line).

### C. Procedure

For all experimental conditions, a three-interval three-alternative forced-choice procedure was used to determine detection thresholds for the signal modulation in the presence of the complex-tone masker modulation. A trial consisted of three intervals separated by 500 ms of silence. In one randomly chosen interval, the carrier was modulated with the sum of the signal and masker modulation. In the two other intervals the carrier was modulated with the masker modulation alone. The signal modulation depth,  $m_s$ , was varied according to a one-up, two-down procedure. This procedure estimates the 70.7 % point on the psychometric function (Levitt, 1971). The step size was 4 dB (in terms of  $20 \log m_s$ ) for the first 2 reversals, was reduced to 2 dB for the following 2 reversals, and was finally set to 1 dB for the remaining 8 reversals that made up one run. The threshold was determined by calculating the median of the signal modulation depth (in dB) at these eight reversals. The procedure was repeated at least three times for each subject and signal configuration. The final threshold estimate was taken as the mean over the last three runs.

## III. RESULTS

### A. Effect of signal phase in two-tone complex masking

Detection thresholds were similar across the three normal-hearing subjects, so mean data are presented here. Figure 4.2 shows detection thresholds and interindividual standard deviations for three different two-tone-complex masker modulations. Masker components at 90 and 120 Hz are indicated by the circles, 97 and 127 Hz by the upward triangles, and 180 and 210 Hz by the squares. The horizontal line indicates the reference thresholds for the 30-Hz signal modulation in the absence of the masker modulation. The downward pointing triangles connected by the dashed lines are mean data from Moore *et*

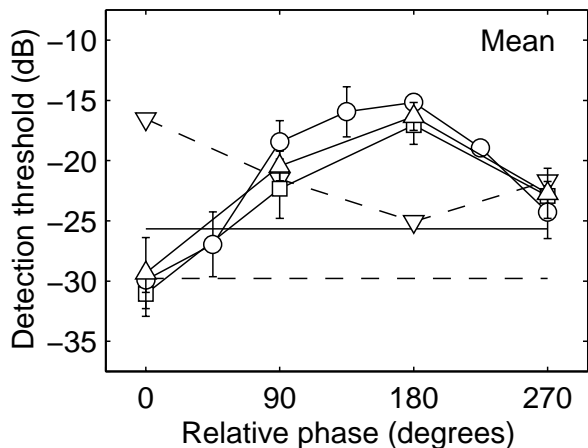


FIG. 4.2: Modulation depth at threshold as a function of the signal starting phase for 30-Hz signal modulation in the presence of a two-tone-complex masker modulation. The horizontal line indicates the modulation depth at threshold in the absence of the modulation masker. Three different pairs masker modulators were used: 90 and 120 Hz (circles), 97 and 127 Hz (upward triangles), and 180 and 210 Hz (squares). The downward pointing triangles connected by the dashed line are mean data from Moore *et al.* (1999), obtained for a 5-Hz signal in the presence of a two-tone complex masker at 40 and 45 Hz.

*al.* (1999), obtained for a 5-Hz signal in the presence of a two-tone complex masker at 40 and 45 Hz. The dashed horizontal line indicates their reference (unmasked) threshold. For the two-tone complex masker at 90 and 120 Hz (circles), thresholds were measured in the range from 0 and 270° in 45° steps. In all other conditions 90° steps were used. For all three conditions, the lowest threshold is observed for a signal phase of 0°, when the maxima in the signal modulation coincided with the maxima in the envelope of the masker. Thresholds for this “in-phase” condition are below the reference threshold. Thresholds increase as the relative phase increases. The maximum threshold is observed for a phase relation of 180°, referred to as the “anti-phase” condition. On average, similar thresholds were obtained for the four signal phases (0, 90, 180, and 270°) that were used for all three different masker modulators. The maximum difference between the threshold curves is about 4 dB (at 90°). The maximum phase effect – the threshold difference between the in-phase and anti-phase condition – is about 15 dB.

The data by Moore *et al.* (1999), indicated by the downward triangles, show a phase effect opposite the that obtained in the present study. The maximum threshold was obtained in the in-phase condition (0°), whereas the minimum threshold was obtained for the anti-phase condition (180°). Masked thresholds were higher than the reference threshold in that study, for all signal-masker phase relations. This reference threshold is about 4 dB lower than that obtained in the present study. The maximum phase effect was about 9 dB. As mentioned above, the data by Moore *et al.* (1999) were obtained for a 5-Hz signal modulation, in contrast to the 30-Hz signal modulation used in the present study. However, the difference in the threshold patterns is not due to the difference in signal frequency. Verhey *et al.* (2002) showed similar masking effects as observed in the present study for additional signal frequencies of 5 and 90 Hz.

## B. Effect of signal phase in three-tone-complex masking

Figure 4.3 shows mean detection thresholds for the normal-hearing subjects and a 30-Hz signal modulation in the presence of a three-tone complex masker. The masker

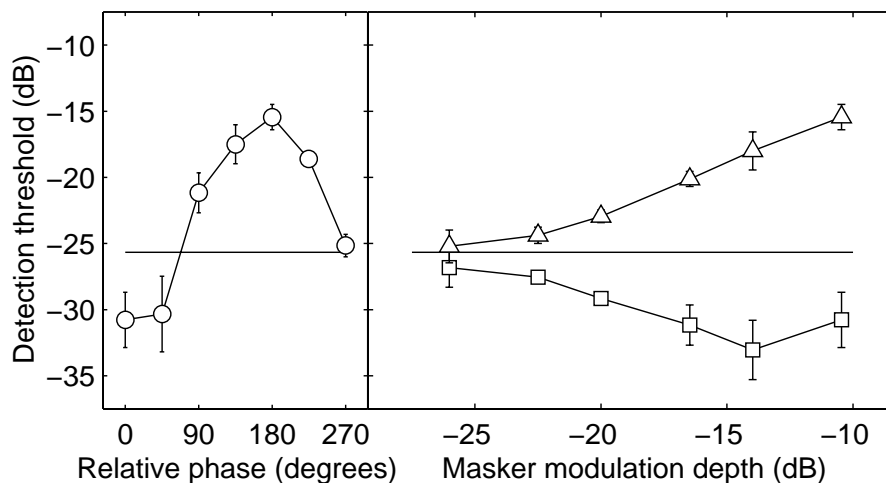


FIG. 4.3: Mean data for the normal-hearing subjects. The left panel shows modulation depth at threshold for a 30-Hz signal modulation in the presence of a three-tone complex masker modulation at 150, 180, and 210 Hz. The right panel shows the threshold as a function of the modulation depth of the highest component of the three-tone complex (for details see text). The triangles indicate data for a signal starting phase of  $180^\circ$ . The circles are for a starting phase of  $0^\circ$ . In both panels, the horizontal line indicates the threshold in the absence of the masker modulation.

consisted of a 180-Hz component that was modulated at 30 Hz. The left panel shows the detection threshold as a function of the signal modulation phase for a masker modulation depth of 0.3. The right panel shows thresholds for the anti-phase (triangles) and the in-phase (squares) conditions, as a function of the masker modulation depth  $m_m$ , in dB. The horizontal line indicates the reference threshold for the 30-Hz signal. The threshold curve as a function of the signal phase (left panel) is very similar to that obtained for the two-tone-complex maskers (Fig. 4.2). Threshold is highest at  $180^\circ$  signal phase (anti-phase condition) and lowest at  $0^\circ$  signal phase (in-phase condition). For all signal phases except  $0^\circ$  and  $45^\circ$ , thresholds are above reference threshold. The maximum phase effect amounts to about 15 dB. The right most data points (at -10 dB) in the right panel of Fig. 4.3 correspond to the maximum and minimum threshold observed in the left panel. For the anti-phase condition (triangles), thresholds continuously decrease with decreasing masker modulation depth. For the in-phase condition (squares), the thresholds decrease between -10 and -14 dB. Below -14 dB, thresholds increase with decreasing masker modulation depth. For decreasing masker modulation depths, the threshold difference between the anti- and the in-phase condition decreases and thresholds asymptote against the reference threshold. A considerable phase effect of 6 dB persists for a masker modulation depth as low as -20 dB, i.e., for a masker modulation depth of the 180-Hz component that is only 6 dB above its threshold.

Figure 4.4 shows mean data comparable to Fig. 4.3 obtained for hearing-impaired listeners. The result for the phase dependent threshold pattern in the left panels is quite clear cut. The data show a distinct threshold minimum and maximum in the in-phase ( $0^\circ$ ) and anti-phase ( $180^\circ$ ) condition, respectively. For a signal phase of  $0^\circ$  the masked

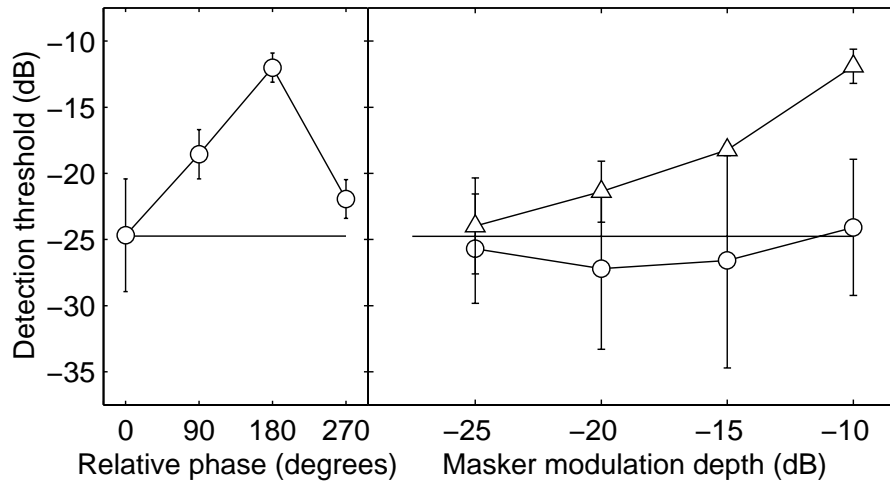


FIG. 4.4: Same as in Fig. 4.3 for hearing-impaired listeners. The upper two panels show individual data, the lower two panels show mean data.

threshold reaches the reference threshold, while in all other conditions masked thresholds are well above the reference threshold. Thresholds as a function of the masker modulation depth (right panel) show more variability across subjects, especially in the in-phase condition (circles). In general, the data show a diminishing threshold difference between the anti-phase (triangles) and in-phase condition with decreasing modulation depth of the masker. The maximum threshold difference of about 14 dB is observed for the maximum modulation depth of the masker. For a modulation depth of -25 dB, masked thresholds are similar to the reference threshold for both phase conditions. The pattern of results for the hearing-impaired subjects is in line with the results obtained for the normal hearing (Fig. 4.3). Thresholds as a function of the signal-masker phase relation and the effect of masker-modulation depth are similar for the two groups of subjects. The main difference between the average data obtained for normal-hearing and hearing-impaired listeners is the amount of “negative” masking, where the threshold in presence of the masker is below the reference threshold, as observed in the in-phase condition. For the normal-hearing listeners negative masking is much more profound and reaches about 8 dB for a signal phase of  $0^\circ$  and a masker-modulation depth of -14 dB (right panel of Fig. 4.3). Negative masking never exceeds 2 dB in the average hearing-impaired data.

#### IV. MODELING THE RESULTS

The data are compared to predictions of a model of AM processing in the auditory system. Three versions of the model are evaluated here differing in the nonlinear processing of the stimulus envelope. The aim of the present study is to assess the extent to which a specific kind of nonlinearity can account for the results. The model does not specify the exact neural realization of the process. The mechanisms in the auditory system may be described by a combination of these simple nonlinearities. This is discussed in Section V..



## A. General model structure

The first stage of all models evaluated here is a linear gammatone filter that simulates the bandpass characteristic of the basilar membrane (Patterson *et al.*, 1987). The center frequency was set equal to the carrier frequency as used in the normal-hearing listeners (5 kHz). Within the model, it is assumed that detection is based on the filter having the greatest signal-to-noise ratio which, in the absence of masking noise, is the on-frequency filter. As shown in Ewert *et al.* (2002), the envelope in off-frequency filters contains a weak envelope component even without nonlinear processing. This component is, however, negligible for the current model predictions. In the second stage, the Hilbert envelope of the signal at the output of the filter is extracted. A model-specific nonlinearity introduces a component at the envelope frequency. The output of the nonlinear stage is then analyzed by a modulation filter centered at the signal-modulation frequency. The stage was used since a number of psychophysical and physiological data indicate the existence of such an envelope-frequency selectivity in the auditory system (see Section 1.2). A Gaussian internal noise was added to the internal envelope representation in order to introduce a lower limit of resolution. The internal noise was assumed to be independent for every sample point and its variance was adjusted to account for the reference (unmasked) threshold for each of the three model versions. It was then kept constant for all model predictions. Cosine-squared rise-fall envelopes of the same length as used for signal generation were applied to the internal representation of the stimuli. The decision device calculated the rms (root-mean square) value of the output of the modulation filter for the three different signals in a trial, i.e., the two reference intervals and the stimulus containing the additional signal modulation. The signal was considered to be detected if the signal interval gave the largest rms value. Only the rms value was considered since Moore and Sek (2000) showed that other measures such as the min-max value or the crest factor are not consistent with their data on the perception of three-component amplitude modulation. Thresholds were predicted using the same adaptive procedure as for the experimental data. The final threshold estimate was averaged across the threshold predictions for ten runs of the adaptive procedure.

## B. Model-specific parameters

Data by Verhey *et al.* (2002), obtained for a broader range of signal frequencies (5 to 90 Hz), suggest that the nonlinear mechanisms can be described as instantaneous mechanisms to a good first approximation. Thus, only instantaneous (fast-acting) nonlinearities are considered here. It is not excluded, however, that the nonlinear mechanisms involved

---

<sup>2</sup>Moore *et al.* (1999) also evaluated a model where a sliding temporal window was used instead of a modulation filter. The temporal window leaves the slow modulations unchanged whereas higher modulations are attenuated. A similar approach was proposed in (Viemeister, 1979). He proposed that the modulation perception of the auditory system can be modeled by assuming a modulation low-pass filter. A low-pass filter approach cannot account for the modulation frequency selectivity as observed in several psychophysical and physiological data. Thus, such an approach is not considered in the present study.

in auditory processing of amplitude modulation are to some degree time-varying as discussed in Moore and Sek (2002). Three different model versions were used. The models differ in the nonlinearity that introduces the envelope component.

i) Compression model: The envelope was compressed by raising it to the power of 0.4. This value was also used by Moore *et al.* (1999).

ii) Threshold model: A constant value equivalent to the mean of the steady-state portion of the internal envelope representation was used as threshold. The maximum between the envelope and this threshold value was calculated. Shofner *et al.* (1996) proposed a threshold as a possible physiologically motivated nonlinearity to account for their data on modulation beating.

iii) Venelope model: The venelope is generated by calculating the envelope of the ac-coupled envelope of the stimuli. The venelope is then multiplied by a factor of 0.3 and added to the original envelope. This procedure was proposed in Ewert *et al.* (2002). The factor of 0.3 was empirically chosen in order to account for the amount of masking observed in the data.

### C. Model predictions

Figure 4.5 shows model predictions (filled symbols) together with the empirical data for normal-hearing listeners replotted from Fig. 4.3 (open symbols). The left panels show predicted thresholds for the 30-Hz signal modulation in the presence of the 3-tone-complex masker as a function of the signal phase. The right panels show predictions for the maximum phase effect as a function of the masker modulation depth. Predictions for the different nonlinear mechanisms integrated in the model are given in the rows of Fig. 4.3 as indicated in the figure. The left panels of Fig. 4.3 show that all models predict thresholds that depend on the signal phase. However, the compression model (upper row) predicts a phase effect opposite to that observed in the data and the other two models. In addition, the model underestimates the phase effect, i.e., the difference between maximum and minimum threshold. Threshold and venelope model show a phase effect that is in line with the empirical data. The threshold model considerably overestimates the maximum phase effect (middle panels of Fig. 4.3). The model predicts a 29-dB effect while the data only show a 15-dB effect. The data in the right panel show that this model overestimates the maximal phase effect for all masker modulation depths. The venelope model (lower panels) provides the best fit to the data. For all masker-modulation depths except for -10 dB, the difference between measured and predicted data is smaller than 2 dB. For -10 dB, the model overestimates negative masking by about 5 dB. This can also be observed in the data as a function of signal-masker phase relation (left lower panel of Fig. 4.3).

## V. DISCUSSION

The experimental results of the present study are consistent with the general finding of Moore *et al.* (1999) that the detection threshold of a signal modulation is affected

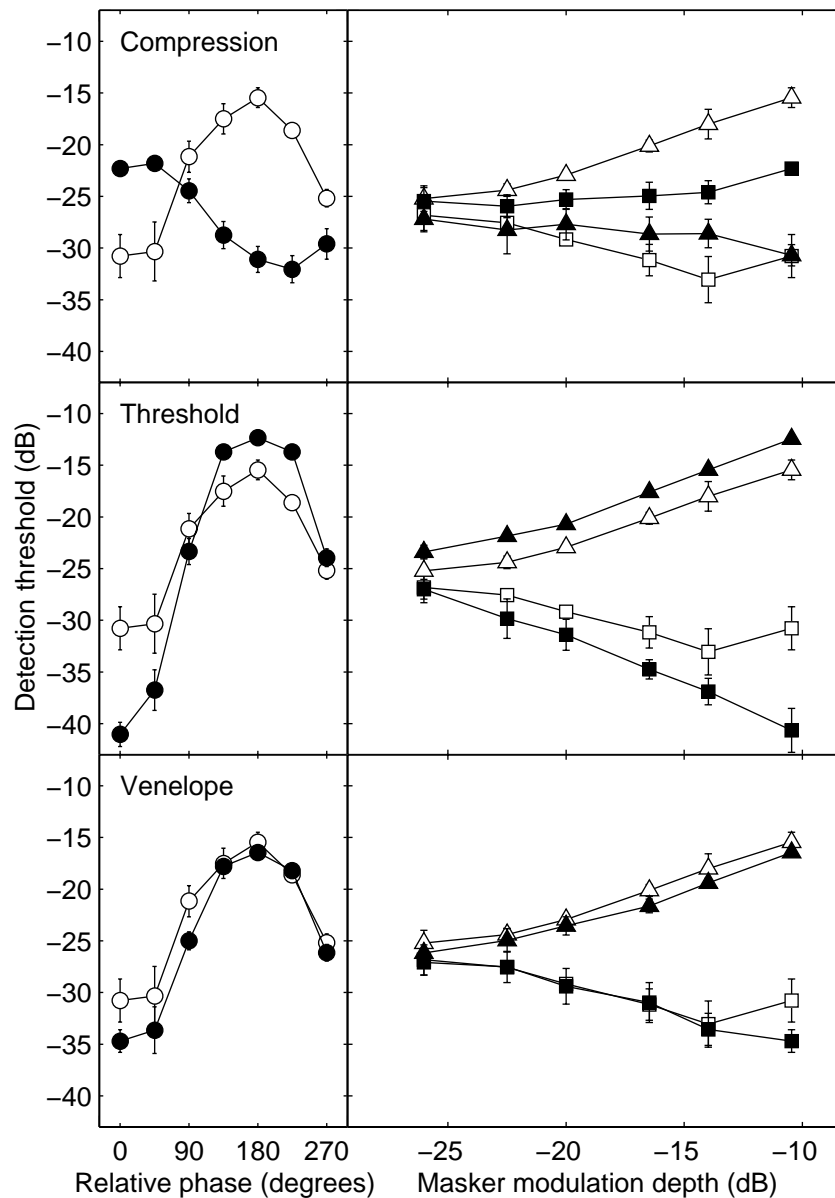


FIG. 4.5: Modulation depth at threshold for a 30-Hz signal modulation in the presence of a three-tone-complex masker modulation (150, 180, and 210 Hz). The left panels show the threshold as a function of the starting phase of the signal. The right panels show the threshold as a function of the modulation depth of the highest component of the three-tone complex masker. Filled symbols indicate model predictions. Open symbols indicate mean empirical data from Fig. 4.3. The top panels show threshold predictions for the model including a compressive nonlinearity. The middle panels show predictions of the threshold model. The bottom panels show threshold predictions obtained with the model which extracts the envelope explicitly.

by the presence of a complex-tone masker beating at the rate of the signal modulation. The results support the hypothesis that a nonlinear mechanism in the auditory system introduces a distortion component at the beat frequency in the internal representation of the envelope. The results could be explained by assuming that the resulting envelope is then processed by a bank of overlapping modulation bandpass filters as proposed by Dau *et al.* (1997a), Dau *et al.* (1999), and Ewert and Dau (2000), where the signal and the

distortion component interact within the filter tuned to the signal frequency. The same magnitude of the phase effect was found irrespective of the spectral distance between signal modulation and masker components. The results in the present study are in agreement with recent studies on the detectability of envelope fluctuations (Lorenzi *et al.*, 2001a,b; Ewert *et al.*, 2002) where it was shown that the detectability of envelope fluctuations is affected by the presence of envelope fluctuations of the same frequency range. The data of these studies and the present study consistently show that the envelope and the envelope interact at some stage of the auditory pathway. The findings are consistent with the idea that both are processed within the same modulation filter.

### **A. Relation to the data from Moore *et al.* (1999)**

The direction of the phase effect in the 5-Hz data of the Moore *et al.* study is opposite to that observed for the 30 Hz data in the present study. Moore *et al.* (1999) obtained the highest detection threshold in the in-phase condition, where the maxima in the signal modulation coincided with the maxima in the masker modulation. In contrast, the threshold maximum in the present study was obtained for the anti-phase condition, and not for the in-phase condition. The findings of the present study are further supported by similar results obtained for 5 and 90-Hz signal modulation frequency in Verhey *et al.* (2002). In summary, it is apparent that all data from the present study consistently show the opposite phase effect to that reported by Moore *et al.* (1999)<sup>3</sup>.

### **B. Effect of the harmonic relation between signal and masker components**

Lorenzi *et al.* (2001b) found noticeable notches in the second-order TMTF when the carrier of the second-order modulation was around twice the envelope frequency. They proposed that this may be connected to a similar effect first observed by Strickland and Viemeister (1996) in the envelope-frequency domain. Strickland and Viemeister (1996) showed that, using one sinusoidal masker modulator, modulation depth at threshold can depend on the relative phase between the signal and the masker if the masker modulation and the signal modulation are harmonically related. The present data for the 30-Hz signal modulation in the presence of a two-tone-complex masker modulation indicate that the same phase effect was found for harmonically related masker components (e.g., 90 and 120 Hz) and masker components with no harmonic relation to the signal modulation (97 and 127 Hz, see Fig. 4.2). Thus, an exact harmonic relation between envelope carrier frequency and signal modulation frequency does not seem to be a crucial prerequisite for a notch. The notches in the data from Lorenzi *et al.* (2001b) may reflect interindividual variation in the performance for different envelope frequencies. Also, the notches were found to be rather broad and not prominent for all subjects.

---

<sup>3</sup>Recently, Moore mentioned in his personal communication that he and his colleagues were not able to replicate their results presented in Moore *et al.* (1999), but instead obtained phase effects similar to those reported here.

### C. Role of compression

Moore *et al.* (1999) showed (for a 5-Hz signal modulation masked by a two-tone complex masker) that a compressive nonlinearity can predict the effect they found for the signal phase relative to the beat cycle of a two-tone-complex masker. They showed that compression causes a threshold increase if the maxima in the masker envelope coincide with the maxima in the signal modulation waveform whereas it predicts a threshold decrease compared to the quiet threshold for an anti-phase signal modulation. Figure 4.5 (upper panels) showed similar predictions for a 30-Hz signal modulation masked by a three-tone-complex masker-modulation using a similar model. The compression model was shown to be able to predict an interaction between the masker envelope and the signal modulation, because the compression introduces a distortion component at the difference frequency. This distortion component has its minima at the positions of the maxima in the masker envelope. The distortion component cancels the signal modulation and hence increases threshold if the signal is in phase with the masker envelope. For an anti-phase signal, the distortion component is in phase with the signal modulation and enhances detectability.

However, the empirical data in the present study clearly show a threshold minimum in the in-phase condition. Thus, the simulated threshold curve obtained for the compression model is  $180^\circ$  out of phase relative to the experimental data. As a consequence, it does not seem to be plausible that the phase effect is connected to a compressive nonlinearity as, for example, observed at the level of the basilar membrane. Moreover, similar data were collected for normal-hearing and hearing impaired subjects. This demonstrates that the phase of the internal envelope component at the envelope frequency is the same in the two groups of subjects, independent of the amount or presence of cochlear compression. This is also consistent with recent findings by Tandetnik *et al.* (2001). They concluded on the basis of their results with hearing-impaired listeners that cochlear damage and the associated loss or reduction of the compressive nonlinearity in the input-output function of the basilar membrane has little effect on the perception of the envelope.

### D. Role of nonlinearities after peripheral compression

The other two model approaches used in the present study follow the ideas from previous studies (Shofner *et al.*, 1996; Ewert *et al.*, 2002). A model that includes a threshold could account for the direction of the phase effect (Fig. 4.5, middle panels). However, the maximum threshold difference produced by phase changes is considerably overestimated for a large masker envelope carrier (middle component of the three-tone complex). A shift of the threshold cannot solve the weakness of this model approach since it would lead to dramatic effects for low envelope-carrier levels (where the the difference between simulated and predicted thresholds are already small) and only negligible effects for high levels.

The envelope model (Ewert *et al.*, 2002) shows the best agreement with the data (Fig. 4.5, bottom panel). In particular, the maximum amount of the phase effect is predicted for almost the whole range of envelope-carrier levels. In this model, the envelope

was extracted separately, attenuated and then added to the envelope prior to the modulation filter bank. The attenuation within the model is responsible for the magnitude of the phase effect. It was set to 0.3 to match the current results. The predictions would be similar to those of the threshold model if the envelope were not attenuated before summation. This indicates that a separate envelope extraction is needed to account for the data. This does not necessarily imply that the envelope has to be extracted by means of a second envelope as used in the model of Ewert *et al.* (2002). A combination of the envelope approach and the threshold approach may be closer to physiology: The envelope might be extracted by means of a threshold as realized in the threshold approach. However, in contrast to the pure “in-line” process described in the threshold approach, an independent pathway might be used to extract the envelope (as realized in the envelope approach). Hypothetically, the addition of the output of this pathway to the original envelope might be realized physiologically in a three-unit circuit: One unit receives excitatory inputs from two cells with a similar best frequency, one unit with a high threshold extracts the envelope and one unit with a lower threshold encodes the complete envelope.

Note that, theoretically, it is also possible to model the phase effect by introducing an expansive nonlinearity. Such a nonlinearity was proposed by, e.g., Zheng and Zhang (1996) to account for their loudness data with cochlear-implant persons. Derleth *et al.* (2001) supposed an expansive nonlinearity as a stage of retro-cochlear processing within the framework of the perception model by Dau *et al.* (1997a,b).

## E. Role of internal noise

The predictions obtained with the envelope model were in good agreement with the empirical data. While the model can account for the amount and the phase dependency of masking, there remain details that are not accurately predicted. Particularly, threshold predictions are symmetric around the quiet threshold while the data show some asymmetry. For the highest masker modulation depth (see bottom left panel of Fig. 4.5), the threshold maximum (for the anti-phase signal) was substantially higher than the quiet threshold, while the threshold minimum (for the in-phase condition) was only slightly lower than the quiet threshold. The current version of the envelope model cannot account for this asymmetry because of the decision stage assumed in the model. Within the model, the signal was detected if the difference between the rms values for the signal-plus-masker representation and the masker-alone representation at the output of the modulation filter at the signal modulation frequency was greater than zero. Detectability was limited by an internal noise with a constant level that was added prior to the calculation of the rms value (see Section IV.). Thus, it is assumed that the auditory system is sensitive to a certain fixed increment irrespective of the excitation level. Note, that this was also assumed for deterministic carriers in the modulation-filterbank model Dau *et al.* (1997a). The model predictions for the phase effect as a function of the masker level shows that this is a reasonable approach for simulating the data for low masker levels (and therefore low envelope amplitudes). With increasing masker level, thresholds decrease for in-phase

signal modulation, except for the highest masker level used in the present study, where the threshold tends to increase again. This increase cannot be accounted for by the current envelope model (see bottom left panel of Fig. 4.5). This effect might be modeled if it is assumed that the internal noise changes from having a constant rms value at low levels to an internal noise with an rms value which is proportional to the excitation level at higher levels. Experiments on modulation-depth discrimination are in agreement with a different sensitivity for low levels (up to about 10 dB above quiet thresholds) than for higher levels (Wakefield and Viemeister, 1990). A smooth transition could be achieved by assuming that the limiting internal noise is a weighted sum of a constant internal noise and a level-dependent (modulation-depth dependent) internal noise as suggested in Ewert and Dau (2002).

## VI. SUMMARY AND CONCLUSIONS

- Thresholds for detecting 30-Hz signal modulation were affected by the presence of complex-tone masker modulations that had a variation in the modulation depth (envelope) at the signal modulation frequency. The thresholds were consistently lower in the in-phase condition, where the maxima in the signal modulation coincide with the maxima in the masker envelope, than in the anti-phase condition. The results were comparable in normal-hearing and sensorineural hearing-impaired listeners.
- The phase effect was independent of the spectral distance between signal and masker in the envelope spectrum. Similar results were obtained for a two-tone complex masker beating at the rate of the signal modulation and a three-tone complex masker with a well-defined sinusoidal envelope.
- For the 30-Hz signal modulation in the presence of the three-tone complex masker modulation, the threshold increase and decrease (for the in- and anti-phase condition, respectively) was approximately symmetric around the unmasked threshold. The observed phase effect persisted for masker modulation components which were only 6 dB above their own threshold.
- A compressive nonlinearity, as proposed by Moore *et al.* (1999), cannot account for the data. Compared to the experimental data, the threshold curve predicted using a compressive nonlinearity shows the opposite phase effect and underestimates the amount of interference. Similar data obtained in normal-hearing and hearing-impaired subject also argues against this hypothesis.
- The phase effect and the amount of interference between the signal modulation and the envelope fluctuation can be accounted for by assuming that the auditory system extracts the masker envelope prior to a modulation filterbank, as proposed in Ewert *et al.* (2002).

## **ACKNOWLEDGMENTS**

We want to thank Neil F. Viemeister, Brian C. J. Moore and an anonymous reviewer for very helpful suggestions and comments on an earlier version of the manuscript. We are grateful to Birger Kollmeier for his support. This work was supported by the Deutsche Forschungsgemeinschaft (DFG), Research Project SPP 1046.



## CHAPTER 5

# MECHANISMS OF ENVELOPE-FREQUENCY SELECTIVITY

### ABSTRACT

A critical experiment is presented that allows to distinguish between a modulation-filterbank analysis, reflecting a limited-resolution spectral decomposition of the stimulus envelope, and a temporal periodicity analysis of the envelope based on the autocorrelation function (ACF). Modulation-detection thresholds were measured in the presence of two different masker waveforms that have the same temporal periodicity but different envelope-spectral content. Masking produced by a squarewave and a sinusoidal masker modulation at a rate of 4 Hz was measured for signal modulations of 16 and 17 Hz. In contrast to the sinusoidal masker, the squarewave masker exhibits additional components at 12, 20, and 28 Hz (odd harmonics) in the envelope spectrum. It was found that the squarewave masker caused 2 to 7 dB more masking than the sinusoidal masker. The results are in line with the expectations from an envelope-power spectrum model of masking [S. D. Ewert and T. Dau, *J. Acoust. Soc. Am.* **108**, 1181-1196 (2000)], realizing a limited-resolution spectral decomposition of the stimulus envelope. The empirical findings can not be expected for an ACF based modeling approach.

### I. INTRODUCTION

In recent studies, human auditory processing of envelope fluctuations has been successfully described in terms of a frequency-selective mechanism, and has been modelled by a modulation filterbank (Dau *et al.*, 1997a,b). The assumption that the auditory system analyzes the temporal envelope of a sound in independent channels tuned to different envelope frequencies is based on empirical observations that either directly motivate or

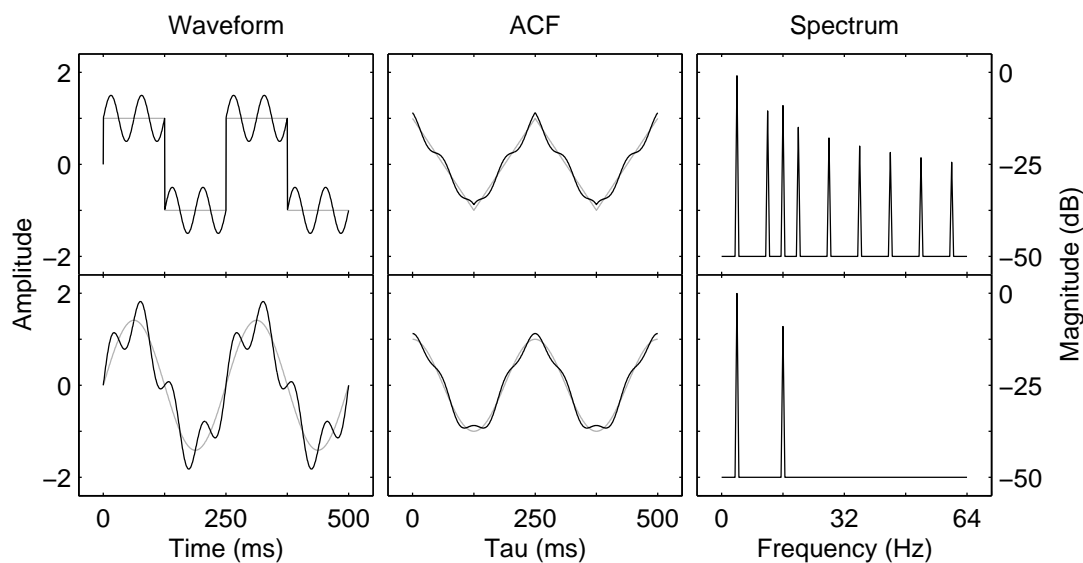


FIG. 5.1: Illustration of two modulation masking paradigms. The left panels display the envelope waveform resulting from a superposition of two modulators at 4 and 16 Hz. The middle and right panels show the autocorrelation function (ACF) and the envelope spectrum of the waveforms, respectively. The upper row of panels shows a superposition of a squarewave modulator at 4 Hz and a sinusoidal modulator at 16 Hz. The lower row shows the same for two sinusoidal modulators at the respective frequencies.

indirectly support this hypothesis. Direct evidence comes from modulation masking experiment (Houtgast, 1989; Bacon and Grantham, 1989; Ewert and Dau, 2000), generally showing tuned masking patterns with maximum masking when the signal- and masker-modulation frequency are close to each other and decreasing amount of masking with increasing spectral separation between the signal and the masker modulation. Also the results of a number of other studies (Fleischer 1982, Dau *et al.*, 1997a,b; Dau *et al.*, 1999; Verhey *et al.*, 1999; Derleth and Dau, 2000) are hardly explained within a common scheme of auditory processing excluding an envelope-frequency selective processing stage.

However, it still remains an open question what mechanisms underly such an envelope-frequency selective process. The present study addresses the question whether the tuning observed in auditory envelope processing reflects a “true” spectral decomposition of the envelope waveform, or whether it reflects a selectivity to repetition periods in the envelope waveform. If the first is true, the interaction of two simultaneously presented modulations at different rates should not only depend on the differences in the envelope repetition periods, but also on the spectral content of the envelope waveforms. In the latter case, only the temporal periodicity of the envelope waveform should be important. Figure 5.1 schematically illustrates two masking conditions that are considered to be useful to address this question. The upper left panel shows the (envelope) waveform resulting from a superposition of a squarewave modulator at 4 Hz and a sinusoidal modulator at 16 Hz (black curve). The lower left panel shows the same for two sinusoidal modulators. The grey curves in the two panels indicate the waveforms of the respective 4-Hz modulators only. The middle and right columns of the figure show the autocorrelation functions (ACFs) and the spectra

of the two waveforms, respectively. The envelope ACF and the envelope power spectrum form a Fourier-transform pair. Intuitively, it may seem plausible that the auditory system is able to process the two superimposed envelope waveforms independently. While the 4-Hz modulation is perceived as a temporal structure or rhythm, the 16-Hz modulation evokes the impression of roughness. This perceptual separation should hold for both stimulus configurations, independent of the exact waveform of the envelope. The ACF (see middle column of Fig. 5.1) is a direct measure of the temporal periodicity prominent in the waveform. Peaks at a certain ACF lag are often used as a measure of the periodicity of the stimulus (envelope) in time-domain models of auditory processing (Meddis and Hewitt, 1991a,b; Meddis and O'Mard, 1997). Expectations for a modulation filterbank (MFB) analysis of the envelope can be derived from the right column of Fig. 5.1. In such a scheme, the detection threshold for the 16-Hz modulation would depend on the integrated envelope power passing the modulation filter at 16 Hz. In case of the squarewave masker modulation at 4 Hz (upper panels), the harmonics in the envelope spectrum would additionally hamper the detection of the 16-Hz signal modulation when compared to the 4-Hz sinusoidal masker modulation (lower panels). Experimental detection thresholds for these two masking conditions are obtained in the following and compared to model predictions based on a modulation-filterbank analysis and an autocorrelation analysis, respectively.

## II. MODULATION MASKING AS A FUNCTION OF THE MASKER WAVEFORM

### A. Method

#### 1. Subjects

Four normal-hearing subjects participated in the study. Their age ranged from 28 to 36 years. All subjects had prior experience in other psychoacoustic experiments. One of the subjects (S1) was author SE, the other subjects were paid an hourly wage for their participation.

#### 2. Apparatus and stimuli

Subjects listened diotically via AKG K501 headphones while seated in a double-walled sound-attenuating booth. Signal generation, presentation and response collection were computer controlled using the AFC software package for MATLAB, developed at the University of Oldenburg. The stimuli were digitally generated at a sampling rate of 48 kHz and converted to analog signals by a two-channel 24-bit DAC including reconstruction filtering (SEKD ADSP 2496). The transfer function of the headphones was digitally equalized (64 point FIR filter) to match a flat amplitude response between 0.1 and 20 kHz, measured with the artificial ear (B&K 4153). A broadband noise and a pure-tone were used as the carrier. The noise carrier ranged from 0 Hz to the Nyquist frequency of 24 kHz.

The pure-tone carrier had a frequency of 4 kHz. A sinusoidal amplitude modulation of 16 or 17 Hz was added to a squarewave or sinusoidal masker modulation of 4 Hz and applied to the carrier. Modulations were applied during the entire carrier duration of 500 ms. The stimuli were gated with 50-ms  $\cos^2$  onset and offset ramps. The basic equation describing the stimuli is as follows

$$s(t) = a \{c(t) [1 + m \sin(2\pi f_s t) + m_m(t)]\} , \quad (5.1)$$

where  $a$  is the amplitude of the stimulus and  $c(t)$  the carrier.  $m$  is the signal-modulation depth and  $f_s$  represents the signal-modulation frequency.  $m_m(t)$  denotes the masker-modulation waveform and was either a sinusoid or a squarewave at 4 Hz. In conditions where a sinusoidal carrier was used in the experiments, the steps in the squarewave function were replaced by 5-ms  $\cos^2$  ramps in order to restrict spectral splatter in the audio-frequency domain. The root-mean-square (rms) modulation depth of the squarewave masker was always -9 dB. In the first experimental series, the sinusoidal masker modulation had the same peak modulation depth as the squarewave masker, and thus a 3 dB lower rms modulation depth of -12 dB. In the second experimental series, both masker waveforms had the same rms modulation depth of -9 dB.

To avoid possible level cues due to the presence of the signal modulation, the stimuli were adjusted to have equal power across the intervals of the forced-choice trial. The overall presentation level was 65 dB SPL.

### 3. Procedure

A three-interval, three-alternative forced-choice paradigm was used to measure modulation-detection thresholds. Subjects had to identify the one randomly chosen interval containing the signal modulation. The two other intervals contained only the masker modulation. A one-up, two-down procedure was used to vary the modulation depth of the signal in dB ( $20 \log m$ ), estimating the 70.7 % correct point of the psychometric function (Levitt, 1971). The three observation intervals were separated by 500 ms of silence. The step size in each run was initially 4 dB and was divided by 2 after every second reversal until it reached 1 dB. At this step size, 6 reversals were obtained and the threshold estimate was calculated as the mean value of  $20 \log m$  at these reversals. Each threshold reported represents the mean of the estimates from at least three runs. On the rare occasions when the standard deviation of the three estimates exceeded 3 dB, an additional estimate was obtained and the first estimate was discarded.

## B. Results and discussion

Figure 5.2 shows results for the first experimental series, where a 16-Hz signal modulation was used. The upper panel shows individual modulation-detection thresholds for the four subjects, indicated by the different symbols. Open symbols represent thresholds in the presence of the squarewave masker modulation, closed symbols are for the conditions

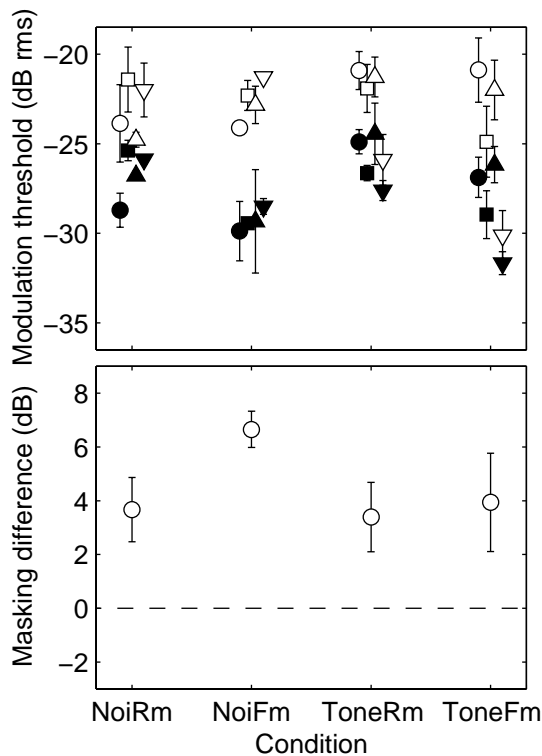


FIG. 5.2: Modulation-detection thresholds (upper panel) and masking difference (lower panel) for a 16-Hz signal modulation in the presence of a squarewave or sinusoidal masker modulation at 4 Hz. The upper panel shows individual thresholds for four subjects indicated by the different symbol types. Open symbols indicate thresholds in the presence of the squarewave masker, closed symbols refer to thresholds in the presence of the sinusoidal masker. The lower panel shows the mean masking difference, computed as the mean of the individual threshold difference between the squarewave and the sinusoidal masker condition. Four combinations of carrier type, broadband noise (Noi) or pure-tone (Tone) and random (Rm) or fixed (Fm) masker phase are indicated by the abbreviations at the abscissa.

where the sinusoidal masker modulation was presented. Thresholds were obtained for a broadband-noise carrier (left most data points) and pure-tone carrier (right most data points), as indicated by the abbreviations Noi and Tone on the abscissa. Rm and Fm stand for a randomly chosen or fixed (zero-degree) starting phase of the masker waveform. The signal modulation always started in sine-phase as expressed in equation 5.1. There is some variation across subjects in the overall sensitivity to amplitude modulation. All subjects show higher thresholds in the presence of the squarewave masker (open symbols) than in the presence of the sinusoidal masker (closed symbols). Except for subject S4 (downward triangles) in the two tonal-carrier conditions and subject S3 (upward triangles) in the NoiRm condition, a difference of about 4 to 5 dB between the two masker types is observed. The mean masking difference between the threshold in the presence of the sinusoidal masker and the threshold in the presence of the squarewave masker is shown in the lower panel of Fig. 5.2. Positive values indicate more masking for the squarewave masker. The maximum masking difference of 6.7 dB is observed in the NoiFm condition, where the signal and masker had a fixed phase relation (see the right panel of Fig. 5.1). In all other conditions, the mean masking difference amounts to about 3 to 4 dB.

In this experiment, the masker waveforms had identical peak values, resulting in a 3 dB higher rms modulation depth for the squarewave masker than for the sinusoidal masker. Thus, the 4-Hz component of the squarewave masker has an about 2 dB higher level than that of the sinusoidal masker. It thus seems plausible that detection threshold in the presence of the squarewave masker is higher than that in the presence of the sinusoidal masker, if it is assumed that thresholds are proportional to the overall envelope power of the masker. This explanation holds independent of the assumption of frequency selec-

tivity in the envelope frequency domain. However, without the assumption of frequency selectivity, much higher thresholds than observed in the data would have been expected. In this case, the presence of the 4-Hz (fundamental) component of the masker should have strongly elevated the threshold above the unmasked threshold of about  $-30$  dB rms for the experimental conditions chosen here (Viemeister, 1979; Eddins, 1993; Kohlrausch *et al.*, 2000). Another expectation could have been that the “local” modulation depth is an important cue for modulation detection. When a faster signal modulation is superimposed on a slower masker modulation, the effective signal-modulation depth (related to the peak-valley ratio) varies along the low-frequency masker-modulation cycle. In the peaks of the masker this effective modulation depth is reduced and in the valleys it is larger than in the absence of the masker modulation. The resulting “negative masking” was discussed earlier in Houtgast (1989). Following this reasoning, lower thresholds should have been expected in the presence of the squarewave masker than in the presence of the sinusoidal masker. Since both maskers were chosen to have an identical peak modulation depth, the squarewave masker provides a longer observation interval with an enlarged effective signal-modulation depth during the negative half-cycle in comparison to the sinusoidal masker. The data in Fig. 5.1, however, clearly contradict this hypothesis. Thresholds were found to be always higher in the presence of the squarewave masker than in the presence of the sinusoidal masker.

In the second experimental series, the rms modulation depth of the two masker waveforms was identical. Conditions were restricted to a random masker-signal phase relation. In addition to the 16-Hz signal modulation, a 17-Hz signal modulation was used in order to avoid a harmonic relationship of the signal and masker. Experimental results are shown in Fig. 5.3. Again, the upper panel displays individual thresholds for the four subjects and the lower panel shows mean masking differences, computed as the difference between thresholds in the presence of the squarewave masker (open symbols in the upper panel) and the sinusoidal masker (closed symbols in the upper panel). The abbreviations on the abscissa indicate N for a broadband-noise carrier and T for a pure-tone carrier. The letter R denotes the phase randomization of the masker waveform, the number indicates the signal-modulation frequency of 16 or 17 Hz. The main observation from the upper panel of Fig. 5.3 is that all subjects show higher thresholds in case of the squarewave masker than in case of the sinusoidal masker. Only subject S4 (downward triangles) shows almost no threshold difference between the two conditions where a pure-tone carrier was used (TR 16 and TR 17). The lower panel summarizes the mean effect of carrier type. On average, 2 to 4 dB more masking is observed for the squarewave than for the sinusoidal masker. A two-way ANOVA (MATLAB R12.1) was conducted using all stimulus configurations (masker type [2] x condition [4]). The effect of masker type was highly significant ( $p < 0.001$ ) for both, the threshold data (upper panel of Fig. 5.3) and the masking-difference data (lower panel). No main effect of condition was found in both cases. Further analysis using a one-way ANOVA (MATLAB R12.1), restricted to each of the four conditions showed a significant effect of masker type for the NR16 ( $p < 0.05$ ) and NR17 threshold data

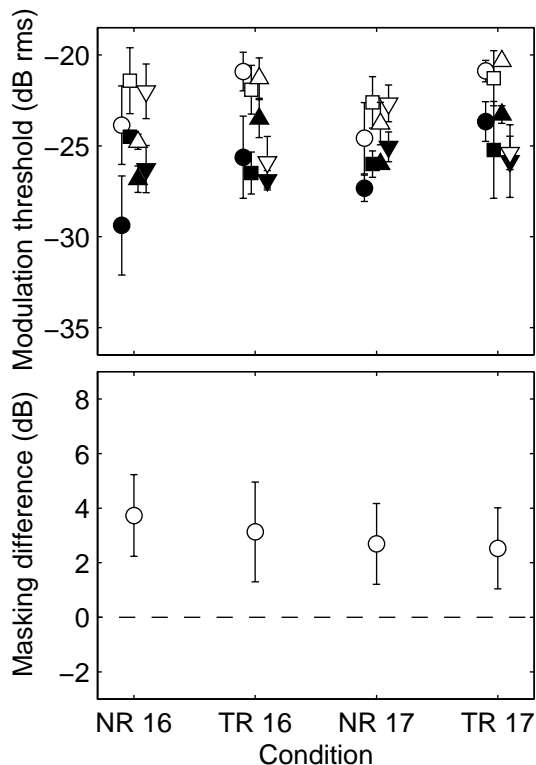


FIG. 5.3: Same as in Fig. 5.2 for different conditions indicated on the abscissa. NR and TR indicate a broadband-noise and pure-tone carrier, respectively. The masker phase was always randomized. A signal-modulation frequency of 16 or 17 Hz was used as indicated by the number. In contrast to the experimental conditions in Fig. 5.2, the square-wave and sinusoidal masker modulation had the same rms modulation depth of -9 dB, whereas they had the same peak modulation depth of -9 dB in Fig. 5.2.

( $p < 0.01$ ). There was a highly significant effect of masker type ( $p < 0.001$ ) in each condition of the masking-difference data. For the individually tested conditions, post hoc comparisons based on Tukey's honestly significant difference (HSD) criterion showed that thresholds for both masker types were significantly different from each other and that the masking difference was significantly different from zero. The masking difference cannot be explained in terms of the overall envelope power of the masker. Since both masker types had the same rms modulation depth, no masking difference would have been expected when disregarding effects of envelope-frequency selectivity. Note that the 4-Hz fundamental of the squarewave masker holds 80% of its waveform power and exhibits about 1 dB less envelope power than the 4-Hz sinusoidal masker in this stimulus configuration. This implies that if only the strength of the periodicity at 4-Hz influenced the detection threshold for the signal modulation, higher thresholds would have been expected in the presence of the sinusoidal masker than in the presence of the squarewave masker. A more quantitative analysis of the data is given in the next section where the data are compared to model predictions.

### III. MODEL PREDICTIONS

#### A. Models

Threshold predictions obtained with two models of envelope processing are presented. The first model is the envelope power spectrum model (EPSM, Ewert and Dau, 2000). The second model is based on the autocorrelation function of the envelope (envelope

autocorrelation model, EACM). The two models are briefly described in the following:

i) The EPSM consists of three processing stages: Hilbert-envelope extraction of the stimuli, a modulation filterbank, and a decision device. The modulation-filterbank design was chosen according to Ewert and Dau (2000). Only the filter tuned to the signal modulation was considered. Effects of peripheral bandpass filtering were not taken into account. The model has a similar structure as Viemeister's (1979) leaky-integrator model, but assumes a modulation filterbank instead of a low-pass filter. In order to investigate the role of the modulation bandpass filters, a second model version using a 64-Hz first-order low-pass filter as suggested by Viemeister (1979), has also been tested. Predicted thresholds are based on the mean integrated envelope power at the output of the modulation filter tuned to the signal-modulation frequency, or the modulation low-pass filter, respectively. It was assumed that the detection of the signal in the presence of the masker modulation requires a 1-dB increase in envelope power. In addition, an absolute lower limit of resolution corresponding to a modulation depth of -30 dB (-33 dB rms) was assumed in case of the pure-tone carrier. For the broadband noise carrier, the lower limit was set to -25 dB (-28 dB rms) based on the modulation detection data in Viemeister (1979).

ii) The EACM performs a Hilbert-envelope extraction, followed by the computation of the autocorrelation function and a decision device. The decision device computes the root-mean-square (rms) distance of the masker-alone and signal-plus-masker output of the model. Detection of the signal requires a certain rms distance in relation to the rms value of the masker-alone model output. The model was calibrated to predict a threshold of -25 dB rms (as observed in the data) for the signal modulation in the presence of the sinusoidal masker modulation applied to the pure-tone carrier (TR 16).

## B. Results and discussion

Figure 5.4 shows model predictions (closed symbols) together with mean empirical data (open symbols) replotted from Fig. 5.3. The upper panel shows modulation-detection thresholds for 16 and 17 Hz in the presence of the sinusoidal 4-Hz masker, indicated by the circles, and in the presence of the squarewave masker, indicated by the squares. Data and EPSM predictions are shown for all combinations of the broadband-noise (N) and pure-tone (T) carrier and the two signal-modulation frequencies, as indicated on the abscissa. The masker-signal phase relation was random in all conditions (indicated as R). The EPSM predictions broadly follow the pattern of results observed in the data. The predictions show a realistic threshold difference between the two masker types and the correct overall sensitivity to amplitude modulation. For the TR 16 condition, hardly any difference between data and predictions is observed. Threshold predictions are always slightly higher for the broadband-noise carrier (N conditions) than for the pure-tone carrier (T conditions). An opposite effect is observed in the data. Predictions for the low-pass version of the EPSM are indicated by the collapsed closed square and circle at about -15 dB. In this case, the threshold is mainly determined by the envelope power caused by the 4-Hz (fundamental) component of the two masker waveforms. No difference between



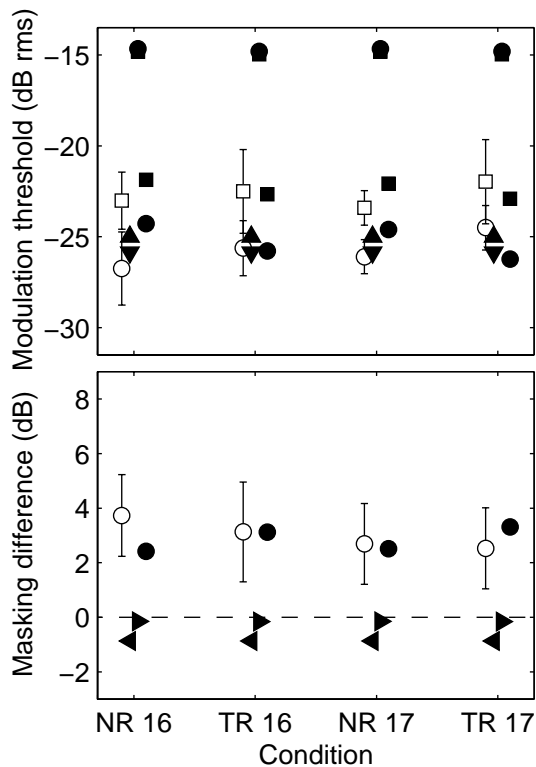


FIG. 5.4: Model predictions (closed symbols) in comparison to mean empirical data (open symbols) replotted from Fig. 5.3. The upper panel shows modulation-detection thresholds and corresponding EPSM predictions in the presence of the square-wave masker (squares) and the sinusoidal masker (circles). The two data points collapsed at -15 dB are predictions for the low-pass filter version of the EPSM. EACM predictions are indicated by the upward and downward triangles for the sinusoidal and squarewave masker, respectively. The lower panel shows the masking difference between both masker types, derived from the upper panel. The circles indicate empirical data and EPSM predictions, the right and left pointing triangles indicate predictions for the low-pass EPSM and the EACM.

the two masker waveforms is predicted.

The closed upward and downward triangle are threshold predictions obtained with the EACM for the sinusoidal and squarewave masker, respectively. The EACM predicts a slightly higher threshold for the sinusoidal masker than for the squarewave masker, in contrast to the data. The overall sensitivity of the model is correct since this was calibrated in the beginning. The model predicts the same thresholds in all experimental conditions.

The lower panel of Fig. 5.4 shows the corresponding masking differences. Positive values indicate higher thresholds for the squarewave masker than for the sinusoidal masker. As expected, the EPSM predictions (closed circles) nicely follow the data (open circles). However, the model predicts a slightly higher masking difference for the pure-tone carrier conditions than for the broadband-noise carrier conditions which is inconsistent with the data. Predictions obtained with the low-pass filter version of the EPSM and the EACM are represented by the right and left pointing triangles, respectively. Both models cannot account for the data.

The model predictions clearly indicate that frequency-selectivity is required in order to account for the data. In agreement with this assumption which is only reflected within the EPSM, the harmonics of the squarewave masker introduce additional masking power in the on-signal modulation filter and lead to higher threshold predictions in the presence of the squarewave masker than in the presence of the sinusoidal masker. This result is in line with the assumption that the auditory system performs a “true” spectral decomposition of the envelope waveform. Threshold predictions obtained for the low-pass version of the EPSM indirectly support the importance of modulation filters. The low-pass version of the EPSM is, indeed, comparable to the EACM, since both models analyze

the broadband envelope. The current EACM would directly correspond to an EPSM version without any modulation filter and assuming a slightly different detection criterion. Since the autocorrelation function and the power spectrum form a Fourier-transform pair, it is clear that a modified EACM could be developed that would be equivalent to the EPSM. Such a model would have to include the convolution of the envelope ACF with the impulse response of the modulation filter, reflecting nothing more than a time-domain implementation of the modulation filter process, whereby phase information would be discarded. As a consequence, there is no simple mechanism based on the autocorrelation function that accounts for the results of the present study. In principle, an envelope-spectrum domain MFB model can be translated into an envelope autocorrelation model. However, such a model would lose the appeal of straight-forward ACF-based models often used in the literature (e.g., Meddis and Hewitt, 1991a,b; Meddis and O'Mard, 1997). ACF models that only use measures like the rms distance of a lag-weighted ACF or the autocorrelation coefficient (peak in the ACF) at a certain time lag, correspond to low-pass or comb-filter type modulation filter functions. They cannot account for a limited-resolution spectral analysis of the envelope as observed in the data.

#### IV. SUMMARY AND CONCLUSIONS

Modulation masking was measured using two different masker modulator waveforms that evoke the same temporal rhythm at a rate of 4 Hz. The signal modulation had a remote rate of 16 Hz. The experiment addressed the question whether frequency selectivity in the envelope frequency domain, as described in a number of earlier studies (Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.* 1997a,b, 1999; Ewert and Dau, 2000), only reflects the existence of channels sensitive to different temporal periodicities in the envelope or whether the auditory system indeed performs a limited-resolution spectral analysis of the envelope waveform. The first hypothesis would suggest a similar processing of a squarewave and a sinusoidal masker modulation when both have the same rate, while the latter hypothesis would suggest effects of the harmonics of the squarewave masker. The results of this study suggest that the auditory system performs a spectral decomposition of complex envelope waveforms. The following conclusions can be drawn:

- Masked thresholds for a 16 or 17-Hz signal modulation are about 3 dB higher in the presence of a squarewave masker modulation than in the presence of a sinusoidal masker modulation at 4 Hz, when both maskers have an identical rms modulation depth.
- The empirical data can be accounted for by the envelope power spectrum model (EPSM), assuming that the threshold for the signal modulation is determined by the amount of envelope power of the masker modulation passing a bandpass modulation filter at the signal frequency. The EPSM predicts that the harmonics of the squarewave masker raise detection thresholds for the signal modulation.

- A modulation low-pass filter version of the EPSM as well as the envelope autocorrelation model (EACM) cannot account for the data. Both models disregard frequency selectivity in the envelope frequency domain. Within both models, thresholds are based on the overall envelope power of the masker and thus only depend on the rms modulation depth of the masker and not on the masker waveform.



# SUMMARY AND CONCLUSIONS

In this thesis, fundamental psychoacoustical experiments on modulation detection and discrimination were conducted to find a better understanding of the basic processing principles of the auditory system when analyzing the temporal envelope of a sound. The empirical data were compared to predictions of models of auditory amplitude-modulation (AM) processing. The questions addressed in the experimental part of each of the chapters were “model driven” and concerned with specific aspects of AM processing. The envelope power spectrum model (EPSM) was developed, which reflects the envelope-frequency analogy to the well known power-spectrum model in the audio-frequency domain (Fletcher, 1940; Patterson and Moore, 1986). The results from each chapter put strong constraints on models of AM processing, including, but not limited to, the EPSM of the current thesis.

In Chapters 1 and 3, the shape and bandwidth of modulation filters were directly estimated using broadband noise and pure-tone carriers. The shape of the masked-threshold patterns measured in both chapters did not vary with the carrier type or the carrier frequency. Within the framework of EPSM, the  $Q$  values of second-order bandpass filters in combination with a fixed first-order low-pass filter at 150 Hz were fitted to the data. AM thresholds were estimated by calculating the (intrinsic) envelope power of the carrier and the masker modulation in the passband of the modulation filter tuned to the signal-modulation frequency. A constant  $Q$ -value of 1 for the bandpass filters accounted well for the masking data.

By measuring AM detection threshold in the presence of two masker modulators that have the same periodicity but different waveforms, Chapter 5 provided further evidence that the auditory system indeed performs a limited-resolution spectral analysis of the envelope, a process which so far has been implicitly assumed within the concept of the modulation filterbank. In contrast, a model that is based on the autocorrelation function of the envelope cannot account for the data. In Chapter 2, it was shown that modulation-detection thresholds and AM-depth discrimination of stochastic and deterministic carriers can be accounted for by the combination of a Weber-fraction-type internal-noise process in combination with the statistical properties of the external stimuli. With this assumption, EPSM was able to accurately account for AM-depth discrimination with broadband-noise and pure-tone carriers as well as for AM detection with bandlimited random- and frozen-noise carriers. Additionally, the predicted thresholds as a function of the carrier bandwidth

indirectly supported the choice of the  $Q$  value for the modulation filter derived in the first chapter. The auditory processing model by Dau *et al.* (1997a,b) was not able to predict these data correctly, since envelope fluctuations are processed essentially linearly within the model and a fixed-variance internal noise is assumed, independent of the internal representation of the envelope. Both, the bandwidth of the bandpass modulation filters ( $Q = 1$ ) and the Weber-fraction-type internal noise process are important changes that have to be tested and incorporated in a future version of the model of “effective” auditory signal processing. In particular, the observation that Weber’s law holds in AM-depth discrimination suggests that the internal representation of the envelope might be transformed by some slow-acting logarithmic non-linearity when the assumption of a fixed-variance internal-noise process is maintained. A first straight-forward approach is given in Appendix C and might be used as a basis for further modeling efforts.

Chapters 3 and 4 addressed the processing of complex modulations. The concept of the “venelope” (defined as the envelope of the ac-coupled envelope) was introduced to generally describe the phenomena originating from beats between the components of complex (multi-component) modulators. In Chapter 3, the role of venelope fluctuations in AM masking paradigms was investigated. It was demonstrated that, in analogy to the “asymmetry of masking” effects observed in the audio-frequency domain, lower masked thresholds can be found when a pure-tone or narrow-band noise signal and a pure-tone masker are used than vice versa. It was shown that listeners use the venelope fluctuations as a detection cue. Thus venelope fluctuations seem to play the same role as envelope fluctuations in corresponding masking paradigms in the audio-frequency domain. Furthermore, the detection of venelope fluctuations can be abolished by the presence of envelope fluctuations of the same rate, indicating that they are not processed independently from each other. The processes underlying the extraction of a venelope component in the auditory system were further investigated in Chapter 4. In contrast to the results from an earlier study (Moore *et al.*, 1999), this chapter demonstrated that the non-linear processing occurring at the stage of the healthy basilar membrane is not responsible for the extraction of the venelope. This was supported by the results obtained with hearing-impaired listeners who showed similar effects as the normal-hearing listeners. Instead, the phase-sensitive interaction of the venelope and the envelope observed in this thesis suggest that either a fast-acting thresholding or expansive non-linearity rather than a fast-acting compressive nonlinearity may be responsible for the extraction of the venelope in the auditory system.

In summary, the modulation masking, detection and discrimination data as well as the processing of complex envelope waveforms could be successfully accounted for on the basis of only a few simple assumptions in the model. The overwhelming result of the present work is that although the envelope of a physical stimulus undergoes a number of complex and not necessarily linear transformations in the auditory system, perception can be satisfactorily modeled to a large extent with regard to the pure physical properties of the stimulus. Nevertheless, a big challenge for future studies will be to incorporate the new knowledge about the mechanisms of envelope processing into more general models

of auditory processing, in order to allow for a combined spectro-temporal analysis in the envelope-frequency as well as in the audio-frequency domain.





# APPENDIX A: Envelope spectra of modulated noise

We consider a modulated waveform  $s(t)$ :

$$s(t) = c(t) (1 + m a(t)) , \quad (\text{A.1})$$

with  $c(t)$  and  $a(t)$  denoting the carrier and the modulator, respectively, and  $m$  representing the modulation depth. As long as the Fourier transform of the modulating waveform  $A(f) = \mathcal{F}\{a(t)\}$  is restricted to frequencies below the lowest frequency component of the carrier spectrum  $C(f) = \mathcal{F}\{c(t)\}$ , the Hilbert envelope  $e_s(t)$  of  $s(t)$  can be expressed as

$$e_s(t) = |\hat{s}(t)| = |\hat{c}(t)| |1 + m a(t)| , \quad (\text{A.2})$$

where  $\hat{s}(t)$  represents the analytic signal of  $s(t)$ . The envelope spectrum  $E_s(f_{env})$  is the Fourier transform of the envelope,  $\mathcal{F}(e_s(t))$ , and can be written as

$$E_s(f_{env}) = E_c(f_{env}) * \mathcal{F}\{|1 + m a(t)|\} , \quad (\text{A.3})$$

where  $E_c(f_{env})$  represents the envelope spectrum of the carrier. Thus, the envelope spectrum of a modulated stimulus can be calculated via convolution of the carrier envelope spectrum and the spectrum of the absolute value of the modulator.

# APPENDIX B: The venelope

Envelope fluctuations are a feature of all waveforms except for pure tones or frequency-modulated tones. In many processing models of the auditory periphery, the envelope is derived from demodulation, realized by (half-wave) rectification followed by lowpass-filtering in order to remove the fine structure of the waveform. The (Hilbert) envelope is less physiologically motivated and reflects a mathematical description. Consider the real waveform  $s(t)$ : The envelope  $he(t)$  of  $s(t)$  is defined as the absolute value of the corresponding analytical signal  $\hat{s}(t) = s(t) + i\tilde{s}(t)$ :

$$he(t) = \sqrt{s^2(t) + \tilde{s}^2(t)}. \quad (\text{B.1})$$

The imaginary part of the analytical signal is the Hilbert-transform of its real part

$$\tilde{s}(t) = \mathcal{H}(s(t)) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{s(t')}{t' - t} dt'. \quad (\text{B.2})$$

Following this notation, the  $n^{\text{th}}$  order (Hilbert) envelope  $he_n(t)$  can be recursively defined by:

$$he_n(t) = \sqrt{(he_{n-1}(t) - \langle he_{n-1}(t) \rangle_t)^2 + \mathcal{H}((he_{n-1}(t) - \langle he_{n-1}(t) \rangle_t))^2)} \quad n \geq 2, \quad (\text{B.3})$$

with  $he_{n-1}(t) - \langle he_{n-1}(t) \rangle_t$  representing the ac-coupled  $(n - 1)^{\text{th}}$  order envelope, where  $\langle \rangle_t$  denotes the temporal long-term average. The second-order envelope ( $n = 2$ ) is referred to as the venelope throughout this study. As a consequence of the diminishing power of higher-order envelopes the second-order envelope might be the only perceptually salient one for acoustic stimuli.

# APPENDIX C: Modifications of the perception model (PEMO)

Two modifications of the perception model (Dau *et al.*, 1997a,b), referred to as PEMO, were tested in conditions of AM-depth discrimination as described in Sec. III. of Chapter 2. The original model design has been modified in order to account for the constant Weber fraction observed in the behavioural AM-depth discrimination data. In the original PEMO, AM-depth discrimination is limited by a fixed-variance internal noise in combination with limitations due to the external variations inherent to the stimulus. The model cannot account for Weber’s law in AM-depth discrimination, since the fixed-variance internal noise leads to constant “absolute” discrimination thresholds rather than constant relative discrimination thresholds. The modifications of the original model design are as follows:

i) The first modification, MOD1, adds an additional internal-noise process to the model. The rms value of the steady-state portion of the internal representation at the output of the modulation filter tuned to the signal-modulation frequency is calculated. A Gaussian noise is generated and bandpass filtered with the respective modulation filter. The rms value of the filtered noise is set to four times the rms value of the internal representation and the two waveforms are added. This procedure adds an Weber-fraction-type internal noise to the internal envelope representation for each of the stimuli during a threshold run.

ii) The second modification, MOD2, applies a logarithmic transformation to the rms value of the internal representation at the output of the modulation filter. As in MOD1, the rms value of the steady-state portion of the internal representation is calculated. The internal representation is then normalized to a rms value of one and multiplied by a factor of  $s = 2(\log \text{rms} + 4)$ . The factor  $s$  is restricted to positive values and is zero for rms values  $\leq 0.0001$ .

Both modified model versions incorporate the second-order modulation-bandpass filters as described in Chapters 1 and 3. In order to maintain the loss of phase-locking to the envelope waveform for modulation frequencies greater than 10 Hz, as it is assumed in the original model, the output of the modulation filter is half-wave rectified and passed through a first-order low-pass filter with a cut-off frequency of 4 Hz. This procedure increasingly reduces information about the temporal structure of the envelope for envelope-frequencies

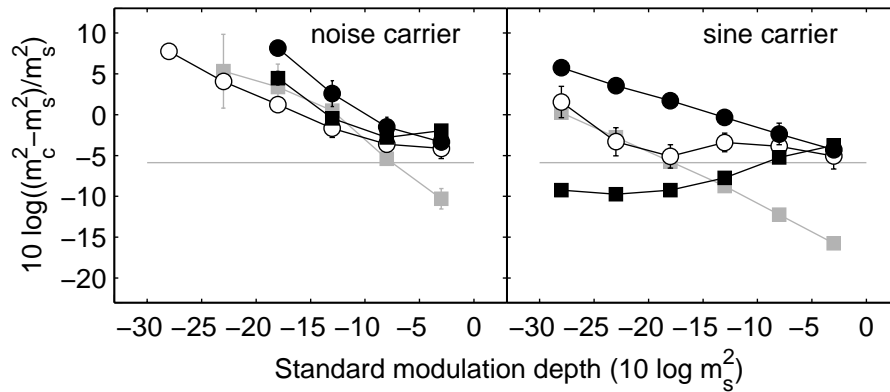


FIG. C.1: Model predictions (closed symbols) and empirical data (open symbols, replotted from Fig. 2.3 in Chapter 2). AM-depth discrimination is shown for a broadband-noise (left panel) and pure-tone carrier. The closed circles represent predictions obtained for the MOD1 model, the closed squares are the MOD2 predictions. Predictions for the original PEMO are indicated by the grey closed squares. For the two lowest standard AM depths ( $-28$  and  $-23$  dB) and the noise carrier (left panel), no predictions could be obtained for both model versions within the upper limit (9 dB) of the tracking procedure.

greater than 4 Hz.

Figure C.1 shows AM-depth discrimination thresholds, plotted as Weber-fractions, for a broadband-noise (left panel) and pure-tone carrier (right panel), comparable to Fig. 2.3 of Chapter 2. Model predictions (black closed symbols), obtained with the two AM-processing models described here, are compared with experimental data (open symbols) and with predictions obtained with the original PEMO (grey closed squares), replotted from Fig. 2.3. The grey line indicates expected thresholds assuming a constant Weber-fraction. The closed circles and squares represent predictions obtained with the modifications MOD1 and MOD2, respectively. Both models show considerable deviations from each other, from the original model, and from the experimental data. For the broadband-noise carrier (left panel), both modifications show realistic Weber-fractions for large AM depths of the standard. With decreasing AM-depths of the standard the models increasingly underestimate detection performance. For the two lowest standard AM-depths ( $-23$  and  $-28$  dB), threshold predictions could not be obtained within the upper limit (9 dB) of the tracking procedure. For standard AM depths of  $-13$  and  $-18$  dB, the predictions obtained with MOD1 (closed circles) are also considerably above those obtained with the original PEMO (grey closed squares). In case of the pure-tone carrier (right panel), MOD1 (closed circles) shows increasing Weber-fractions with decreasing AM depth of the standard, thus underestimating listeners performance. In contrast, MOD2 shows essentially decreasing Weber-fractions and generally overestimates listeners' performance.

In principle, the pattern of results obtained with the two model versions, follows the expectations from the respective modifications. In contrast to the much more simple and linear envelope power spectrum model (see Chapter 2), however, the more complex, non-linear processing within the PEMO additionally influences the model's behaviour. The logarithmic transformation of the rms value of the internal representation within MOD2

should lead to constant Weber-fractions in case of the pure-tone carrier. This is only the case for small standard AM-depths (closed squares in the right panel of Fig. C.1). In order to account for the region of rising Weber-fractions, as observed for standard AM depths smaller than  $-18$  dB in the data, it appears that an additional fixed-variance internal noise would have to be introduced prior to the logarithmic transformation of the internal representation. For the other model, MOD1, the combination of a Weber-fraction-type and a fixed-variance internal noise leads to a slope of the predicted Weber-fraction function intermediate to the original PEMO and the constant Weber-fraction expectation for the pure-tone carrier. The model might be further improved by an adjustment of the two noise sources. The main observation from the predictions for the broadband-noise carrier (left panel) is that both models considerably underestimate listeners' performance for small AM-depths of the standard where the external signal variability of the stimuli limits the performance in the original PEMO. In particular, the additional signal-dependent internal noise introduced in MOD1 overemphasizes the influence of external signal variability.

In conclusion, the two straight-forward "ad hoc" modifications presented here do not allow for correctly predicting the data. A more sophisticated modeling approach is needed to predict AM-depth discrimination within the framework of the PEMO.



# REFERENCES

- Bacon, S.P., and Viemeister, N.F. (1985). “Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners,” *Audiology* 24, 117–134.
- Bacon, S. P., and Grantham, D. W. (1989). “Modulation masking: Effects of modulation frequency, depth, and phase,” *J. Acoust. Soc. Am.* **85**, 2575–2580.
- Bacon, S. P., and Gleitman (1992). “Modulation detection in subjects with relatively flat hearing losses,” *JSHR* 35 **35**, 642–653.
- Bos, C. E., and deBoer, E. (1966). “Masking and discrimination,” *J. Acoust. Soc. Am.* **82**, 708–715.
- Chistovich, L. A. (1957). “Frequency characteristics of masking effect,” *Biophys.* **2**, 743–755.
- 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. (1997a). “Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers,” *J. Acoust. Soc. Am.* **102**, 2892–2905.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). “Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection,” *J. Acoust. Soc. Am.* **102**, 2906–2919.
- Dau, T., Verhey, J. L., and Kohlrausch, A. (1999). “Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers,” *J. Acoust. Soc. Am.* **106**, 2752–2760.
- Derleth, R. P., and Dau, T. (2000). “On the role of envelope fluctuation processing in spectral masking,” *J. Acoust. Soc. Am.* **108**, 285–296.
- Derleth, R. P., Dau, T., and Kollmeier, B. (2000). “Amplitude modulation detection in normal and hearing-impaired listeners,” *subm. to J. Acoust. Soc. Am.* .
- Derleth, R. P., Dau, T., Kollmeier, B. (2001). “Modeling temporal and compressive properties of the normal and impaired auditory system,” *Hear Res.* **159**, 132–149.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). “Effect of temporal envelope smearing on speech reception,” *J. Acoust. Soc. Am.* **95**, 1053-1064.

- Eddins, D. A. (1993). "Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region," *J. Acoust. Soc. Am.* **93**, 470–479.
- Eddins, D. A. (1999). "Amplitude-modulation detection at low- and high-audio frequencies," *J. Acoust. Soc. Am.* **105**, 829–837.
- Egan, J. P., and Hake, H. W. (1950). "On the masking pattern of a simple auditory stimulus," *J. Acoust. Soc. Am.* **22**, 622–630.
- Ehmer, R. H. (1959). "Masking patterns of tones," *J. Acoust. Soc. Am.* **31**, 1115–1120.
- Ewert, S. D., and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* **108**, 1181–1196.
- Ewert, S. D., and Dau, T. (2002). "Internal and external limitations in amplitude-modulation processing," *J. Acoust. Soc. Am.*, in review.
- Ewert, S. D., Verhey, J. L., and Dau, T. (2002). "Spectro-temporal processing in the envelope-frequency domain," *J. Acoust. Soc. Am.*, **112**, 2921–2931.
- Fassel, R. (1995). *Experimente und Simulationsrechnungen zur Wahrnehmung von Amplitudenmodulationen im menschlichen Gehör* Doctoral Thesis, Universität Göttingen.
- Fleischer, H. (1980). "Subjective Größe von Unterschieden im Amplitudenmodulationsgrad von Sinustönen," *Acustica* **46**, 31–37.
- Fleischer, H. (1982). "Modulationsschwellen von Schmalbandrauschen," *Acustica* **51**, 154–161.
- Fleischer, H. (1983). "Modulation thresholds of narrow noise bands". In *Proceedings of the 11th ICA*, 99–102, Paris.
- Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.* **12**, 47–65.
- Fletcher, H., and Munson, W. A. (1937). "Relation between loudness and masking," *J. Acoust. Soc. Am.* **9**, 1–10.
- Formby, C. C. (1987). "Modulation threshold functions for chronically impaired maniere patients," *Audiology* **26**, 89–102.
- 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.
- Florentine, M. (1983). "Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing," *J. Acoust. Soc. Am.* **74**, 1375–1379.
- Florentine, M., Buus, S., and Maason, C.R. (1987). "Level discrimination as a function of level for tones from 0.25 to 16 kHz," *J. Acoust. Soc. Am.* **81**, 1528–1541.
- Glasberg, B.R., and Moore, B.C.J. (1992). "Effects of envelope fluctuations on gap detection," *Hear. Res.* **64**, 81–92.
- 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 discrimination of narrow-band sounds," *J. Acoust. Soc. Am.* **92**, 2586–2597.
- Greenberg, S., and Arai, T. (1998). "Speech intelligibility is highly tolerant of cross-channel spectral asynchrony," *J. Acoust. Soc. Am.* **103**, 3057.



- Hall, J. L. (1997). "Asymmetry of masking revisited: Generalization of masker and probe bandwidth," *J. Acoust. Soc. Am.* **101**, 1023-1033.
- Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **76**, 50-56.
- Hall, J. W., Grose, J. H., and Haggard, M. P. (1989). "Effects of flanking band proximity, number, and modulation pattern on comodulation masking release.," *J. Acoust. Soc. Am.* **87**, 269-283.
- Hall, J. W., and Grose, J. H. (1991). "Some effects of auditory grouping factors on modulation detection interference (MDI)," *J. Acoust. Soc. Am.* **90**, 3028-3035.
- Hartmann, W.M., and Pumplin, J. (1988). "Noise power fluctuation and the masking of sine signals," *J. Acoust. Soc. Am.* **83**, 2277-2289.
- Hartmann, W.M., and Pumplin, J. (1991). "Periodic signals with minimal power fluctuations," *J. Acoust. Soc. Am.* **90**, 1310-1317.
- Hellman, R. P., (1972). "Asymmetry of masking between noise and pure tone", *Perception and Psychophysics* **11**, 241-246.
- Henning, G.B., and Psotka, J. (1969). "Effect of duration on amplitude discrimination in noise," *J. Acoust. Soc. Am.* **45**, 1008-1013.
- Houtgast, T. (1989). "Frequency selectivity in amplitude-modulation detection," *J. Acoust. Soc. Am.* **85**, 1676-1680.
- Jesteadt, W., Wier, C.C., and Green, D.M. (1977). "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 169-177.
- 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.
- Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (1965). "Discharge Patterns of Single Fibers in the Cat's Auditory Nerve," (Research Monographs No. 35) MIT Press, Cambridge, Mass.
- Kohlrausch, A., Fassel, R., and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," *J. Acoust. Soc. Am.* **108**, 723-734.
- Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Püschel, D. (1997). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," *Acustica united with acta acustica* **83**, 659-669.
- Langner, G., and Schreiner, C. (1988). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism," *J. Neurophysiol.* **60**, 1799-1822.
- Lawson, J. L., and Uhlenbeck, G. E. (1950). *Threshold Signals*, volume 24 of *Radiation Laboratory Series*, McGraw Hill, New York.
- Lee, J., and Bacon, S.P. (1997). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration," *J. Acoust. Soc. Am.* **101**, 3688-3693.
- Levitt, H. (1971). "Transformed up-down procedures in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477.
- Lorenzi, C., Micheyl, C., Berthommier, F., and Portalier, S. (1997). "Modulation masking with sensorineural hearing loss," *J. Speech Lang. Hear. Res.* **40**(1), 200-207.

- Lorenzi, C., Simpson, M. I. G., Millman, R. E., Griffiths, T. D., Woods, W. P., Rees, A., and Green, G. G. R. (2001b). "Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers," *J. Acoust. Soc. Am.* **110**, 2470–2478.
- Lorenzi, C., Soares, C., Vonner, T. (2001a). "Second-order temporal modulation transfer functions," *J. Acoust. Soc. Am.* **110**, 1030–1038.
- Martens, J. (1982). "A new theory of multitone masking," *J. Acoust. Soc. Am.* **72**, 397–405.
- Maiwald, D. (1967). "The calculation of modulation thresholds with a model," *Acustica* **18**, 193–207.
- McFadden, D. M. (1986). "Comodulation masking release: Effects of varying level, duration, and time delay of the cue band," *J. Acoust. Soc. Am.* **80**, 1658–1667.
- Meddis, R., and Hewitt, M. J. (1991a). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," *J. Acoust. Soc. Am.* **89**, 2866–2882.
- Meddis, R., and Hewitt, M. J. (1991b). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Phase sensitivity," *J. Acoust. Soc. Am.* **89**, 2883–2894.
- Meddis, R., and O'Mard, L. A. (1997). "A unitary model of pitch perception," *J. Acoust. Soc. Am.* **102**, 1811–1820.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- 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. J. Moore (Academic, London), 251–308.
- Moore, B. C. J., and Sek, A. (2000). "Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation," *J. Acoust. Soc. Am.* **108**, 2337–2344.
- Moore, B. C. J. (1995). *Perceptual Consequences of Cochlear Damage*, (Oxford University Press, Oxford).
- Moore, B. C. J., Alcántara, J. L., and Dau, T. (1998). "Masking patterns for sinusoidal and narrow-band noise maskers," *J. Acoust. Soc. Am.* **104**, 1023–1038.
- Moore, B. C. J., Sek, A., and Glasberg, B. R. (1999). "Modulation masking produced by beating modulators," *J. Acoust. Soc. Am.* **106**, 938–945.
- Oxenham, A. J., and Moore, B. C. J. (1995). "Additivity of masking in normally hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **98**, 1921–1934.
- Oxenham, A. J., and Plack, C. J. (1997). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **101**, 3666–3675.
- Oxenham, A., and Dau, T. (2001). "Modulation detection interference: Effects of concurrent and sequential stream segregation," *J. Acoust. Soc. Am.* **110**, 404–408.
- Ozimek, E., and Sek, A. (1988). "AM difference limens for noise bands," *Acustica* **66**, 153–160.
- Palmer, A. R., (1995). "Neural signal processing," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

- 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).
- 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 Modeling at RSRE, 14–15 December.
- Plack, C. J., and Oxenham, A. J. (1998). "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.* **103**, 1598–1608.
- Püschel, D. (1988). *Prinzipien der zeitlichen Analyse beim Hören* Doctoral Thesis, Universität Göttingen.
- Plomp, R. (1988). "The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function," *J. Acoust. Soc. Am.* **83**, 2322–2327.
- Pumplin, J. (1985). "Low-noise noise," *J. Acoust. Soc. Am.* **78**, 100–104.
- Raab, D.H., and Goldberg, I.A. (1975). "Auditory intensity discrimination with bursts of reproducible noise," *Phys. Rev.* **31**, 867–875.
- Richards, V.M. (1992). "The detectability of a tone added to narrow bands of equal-energy noise," *J. Acoust. Soc. Am.* **91**, 3424–3435.
- Riesz, R.R. (1928). "Differential intensity sensitivity of the ear for pure tones," *J. Acoust. Soc. Am.* **57**, 437–447.
- Rodenburg, M. (1972). *Sensitivity of the auditory system to differences in intensity* Doctoral Thesis, Medical Faculty of Rotterdam.
- Rodenburg, M. (1977). "Investigation 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.
- Rodenburg, M., Verschuure, J., and Brocaar, M. P. (1974). "Comparison of two masking methods," *Acustica* **31**, 99–106.
- Ruggero, M. A., and Rich, N. C. (1991). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," *J. Neurosci.* **11**, 1057–1067.
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). "Basilar membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* **101**, 2151–2163.
- Schulze, H., and Langner, G. (1997). "Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm," *J. Comp. Physiol. A* **181**, 651–663.
- Schulze, H., and Langner, G. (1999). "Auditory cortical responses to amplitude modulation with spectra above frequency receptive fields: evidence for a wide spectral integration," *J. Comp. Physiol. A* **185**, 493–508.
- Sek, A., and Moore, B. C. J. (2002). "Mechanisms of modulation gap detection," *J. Acoust. Soc. Am.* **111**, 2783–2792.
- Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131–141.

- Sheft, S., and Yost, W. A. (1990). "Temporal integration in amplitude modulation detection," *J. Acoust. Soc. Am.* **88**, 796–805.
- Sheft, S., Yost, W. A. (1997). "Modulation detection interference with two-component masker modulators," *J. Acoust. Soc. Am.* **102**, 1106–1112.
- Shofner, W. P., Sheft, S., and Guzman, S. J. (1996). "Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes," *J. Acoust. Soc. Am.* **99**, 3592–3605.
- Strickland, E. A., Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**, 3638–3646.
- Strickland, E. A., Viemeister, N. F. (1997). "The effects of frequency region and bandwidth on the temporal modulation transfer function," *J. Acoust. Soc. Am.* **102**, 1799–1810.
- Strickland, E. A. (1999). "The effects of frequency region and level on the temporal modulation transfer function," *J. Acoust. Soc. Am.* **107**, 942–952.
- Tandetnik, S., Garnier, S., Lorenzi, C. (2001). "Measurement of first- and second-order modulation detection thresholds in listeners with cochlear hearing loss," *Br. J. Audiol.* **35**, 355–364.
- Takahashi, G.A., and Bacon, S.P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Lang. Hear. Res.* **35**(6), 1410–1421.
- Tanner, W. P., and Sorkin, R. D. (1972). "The theory of signal detectability," in *Foundations of modern auditory theory*, edited by J. V. Tobias (Acad. Press, N.Y.).
- van den Brink, W. A. C., and Houtgast, T. (1990a). "Efficient across-frequency integration in short-signal detection," *J. Acoust. Soc. Am.* **87**, 284–291.
- van den Brink, W. A. C., and Houtgast, T. (1990b). "Spectro-temporal integration in signal detection," *J. Acoust. Soc. Am.* **88**, 1703–1711.
- Verhey, J.L. (2002). "Modeling the influence of inherent amplitude fluctuation simultaneous masking experiments," *J. Acoust. Soc. Am.* **111**, 1018–1025.
- Verhey, J.L. and Dau, T. and Kollmeier, B. (1999). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," *J. Acoust. Soc. Am.* **106**, 2733–2745.
- Verhey, J. L., Ewert, S. D., and Dau, T. (2002). "Modulation masking produced by complex tone modulators," *subm. to J. Acoust. Soc. Am.* .
- Viemeister, N. V. (1977). "Temporal factors in audition: A systems analysis approach," in *Psychoacoustics 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.
- Wakefield, G.H., and Viemeister, N.F. (1990). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," *J. Acoust. Soc. Am.* **88**, 1367–1373.
- Wegel, R. L., and Lane, C. E. (1924). "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear," *Phys. Rev.* **23**, 266–285.

- Yates, G. K. (1990). "Basilar-membrane nonlinearity and its influence on auditory nerve rate-intensity functions," *Hear. Res.* **45**, 203–220.
- Yost, W. A., Sheft, S. (1988). "Detecting amplitude modulation of sinusoidal carriers," *J. Acoust. Soc. Am. Suppl.* **1 83**, S35.
- Yost, W. A., 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.
- Yost, W. A., Sheft, S. (1997). "Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions," *Auditory Neuroscience* **3**, 401–414.
- Zwicker, E. (1956). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs," *Acustica* **6**, 356–381.
- Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). "Critical bandwidth in loudness summation," *J. Acoust. Soc. Am.* **29**, 548–557.
- Zwicker, E., and Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger*, (Hirzel Verlag, Stuttgart).



# CURRICULUM VITAE

<b>Name</b>	Stephan Ewert
Birth	05.07.1972, Oldenburg, Germany
Marital status	Single
Nationality	German
<b>Education</b>	
1979-1983	Elementary school (Grundschule Staakenweg) in Oldenburg, Germany.
1983-1992	Grammar school (Orientierungsstufe Marschweg, Altes Gymnasium Oldenburg) in Oldenburg, Germany. Final examination (Allgemeine Hochschulreife) in June, 1992
10/92-9/93	Selective service
10/93-5/99	Study of Physics at the University of Oldenburg, Germany.
10/97	First graduation in Physics (Vordiplom).
5/99	Master's degree in Physics (Diplom-Physiker). Degree class I. Title: "Frequency selectivity in auditory processing of amplitude modulation".
since 5/1999	Work on doctoral thesis. Title: "Auditory spectro-temporal processing in the envelope-frequency domain: Experiments and models".
3/2001-6/2001	Visiting Scientist at the Research Lab of Electronics at the Massachusetts Institute of Technology, Cambridge, MA, USA.
<b>Experience</b>	
6/1999-4/2002	Position as "Wissenschaftlicher Angestellter" in the DFG Special Research Project (SFB 517) "Neural Basis of Cognitive Function" at the University of Oldenburg.
since 5/2000	Associate member of the "European Graduate School for Neurosensory Science, Systems and Applications" at the University of Oldenburg supported by the Deutsche Forschungsgemeinschaft.
since 5/2002	Position as "Wissenschaftlicher Angestellter" in the DFG Research Project (SPP 1046) "Zeitgebundene Verarbeitung im zentralen auditorischen System" at the University of Oldenburg.





# DANKSAGUNG

An dieser Stelle möchte ich Prof. Dr. Dr. Birger Kollmeier für die Ermöglichung dieser Arbeit in der “Medizinischen Physik” an der Universität Oldenburg danken. Unter seiner Leitung konnte diese Arbeit in einer hervorragenden, vertrauensvollen Arbeitsatmosphäre mit großen Freiheiten entstehen.

Prof. Dr. Volker Mellert danke ich für sein Interesse und die Übernahme des Korreferates. Mein besonderer Dank gilt Dr. Torsten Dau (“T blank Dau”) für die erstklassige Betreuung, die zahlreichen, unendlich langen Diskussionen und die niemals versiegende Unterstützung und Motivation, die ich von ihm während der gesamten Zeit erfahren habe.

Oliver Wegner danke ich für seine Hilfsbereitschaft bei diversen, trivialen oder komplizierten Fragen. Die hervorragende Arbeitsatmosphäre in unserem gemeinsamen Büro und sein guter Rat haben zum Gelingen dieser Arbeit erheblich beigetragen.

Dr. Jesko L. Verhey möchte ich für die kritischen Auseinandersetzungen und Diskussionen danken.

Allen Mitgliedern der “Medizinischen Physik” und insbesondere den Mitgliedern des APO- und NEURO- “Bläschens” sei hier noch für die stets lockere und angenehme Arbeitsatmosphäre gedankt. Insbesondere Dr. Michael Bellmann und Jörg Damaschke danke ich für die unermüdliche Teilnahme an meinen niemals endenden Messreihen. Dr. Birgitta Gabriel, Müge Kaya und Kerstin Sommer danke ich für die Unterstützung bei den Messungen mit Schwerhörenden.

Bei Prof. Dr. Armin Kohlrausch möchte ich mich für sein Interesse und seine Anregungen zu dieser Arbeit, sowie bei Dr. Andrew Oxenham für die schöne Zeit in Cambridge, MA, USA bedanken.

An dieser Stelle möchte ich auch allen Menschen danken, die mich während meiner langen “nächste Woche gebe ich ab” und “morgen kommen sie ab” Zeit ertragen und unterstützt haben.



## **Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst habe und nur die angegebenen Hilfsmittel verwendet habe.

Oldenburg, den 07. November 2002

Stephan Ewert

