Open access • Journal Article • DOI:10.1121/1.420344

Modeling auditory processing of amplitude modulation I. Detection and masking with narrow-band carriers

Torsten Dau, Birger Kollmeier, AG Armin Kohlrausch

Institutions: University of Oldenburg

Published on: 01 Nov 1997 - Journal of the Acoustical Society of America (Acoustical Society of America)

Topics: Amplitude modulation and Pulse-frequency modulation

Modeling auditory processing of amplitude modulation I. Detection and masking with narrow-band carriers

Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers[a)]

Torsten Dau[b)] and Birger Kollmeier
Carl von Ossietzky Universität Oldenburg, Graduiertenkolleg Psychoakustik, AG Medizinische Physik, D-26111 Oldenburg, Germany

Armin Kohlrausch
IPO Center for Research on User-System Interaction, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 28 June 1996; accepted for publication 4 August 1997)

This paper presents a quantitative model for describing data from modulation-detection and modulation-masking experiments, which extends the model of the ''effective'' signal processing of the auditory system described in Dau *et al.* [J. Acoust. Soc. Am. **99**, 3615–3622 (1996)]. The new element in the present model is a modulation filterbank, which exhibits two domains with different scaling. In the range 0–10 Hz, the modulation filters have a constant bandwidth of 5 Hz. Between 10 Hz and 1000 Hz a logarithmic scaling with a constant Q value of 2 was assumed. To preclude spectral effects in temporal processing, measurements and corresponding simulations were performed with stochastic narrow-band noise carriers at a high center frequency (5 kHz). For conditions in which the modulation rate (f_{mod}) was smaller than half the bandwidth of the carrier (Δf), the model accounts for the low-pass characteristic in the threshold functions [e.g., Viemeister, J. Acoust. Soc. Am. **66**, 1364–1380 (1979)]. In conditions with $f_{\mathrm{mod}} > \Delta f/2$, the model can account for the high-pass characteristic in the threshold function. In a further experiment, a classical masking paradigm for investigating frequency selectivity was adopted and translated to the modulation-frequency domain. Masked thresholds for sinusoidal test modulation in the presence of a competing modulation masker were measured and simulated as a function of the test modulation rate. In all cases, the model describes the experimental data to within a few dB. It is proposed that the typical low-pass characteristic of the temporal modulation transfer function observed with wide-band noise carriers is not due to ''sluggishness'' in the auditory system, but can instead be understood in terms of the interaction between modulation filters and the inherent fluctuations in the carrier. [S0001-4966(97)05611-7]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [JWH]

INTRODUCTION

Temporal resolution in the auditory system, or the ability to resolve dynamic acoustic cues, is very important for the processing of complex sounds. A general psychoacoustical approach to describing temporal resolution is to measure the threshold for detecting changes in the amplitude of a sound as a function of the rate of the changes. The function which relates threshold to modulation rate is called the temporal modulation transfer function (TMTF) (Viemeister, 1979). The TMTF might provide important information about the processing of temporal envelopes. Since the modulation of a sound modifies its spectrum, wide-band noise is often used as a carrier signal in order to prevent subjects from using changes in the overall spectrum as a detection cue; modulation of white noise does not change its long-term spectrum (e.g., Burns and Viemeister, 1981). The subject's sensitivity for detecting sinusoidal amplitude modulation of a broadband noise carrier is high for low modulation rates and decreases at high modulation rates. It is therefore often argued that the auditory system is ''sluggish'' in following fast temporal envelope fluctuations. Since this sensitivity to modulation resembles the transfer function of a simple low-pass filter, the attenuation characteristic is often interpreted as the *low-pass characteristic* of the auditory system. This view is reflected in the structure of a widely accepted model for describing the TMTF (Viemeister, 1979).

It is often argued that the auditory filters play a role in limiting temporal resolution (e.g., Moore and Glasberg, 1986), especially at frequencies below 1 kHz where the bandwidths of the auditory filters are relatively narrow, leading to longer impulse responses or ''ringing'' of the filters. However, the response of auditory filters at high center frequencies is too fast to be a limiting factor in most tasks of temporal resolution (Ronken, 1970; Green, 1973). Thus there must be a process at a more central level of the auditory system than the peripheral auditory filters which limits temporal resolution and causes the ''sluggishness'' in following fast modulations of the stimulus envelope.

Results from several studies concerning modulation masking, however, are not consistent with the idea of only one broad filter, as reflected in the TMTF. Houtgast (1989) designed experiments to estimate the degree of frequency selectivity in the perception of simultaneously presented am-

[a)]Part of this research was presented at the 129th meeting of the Acoustical Society of America [T. Dau, B. Kollmeier and A. Kohlrausch, ''Modeling modulation perception: modulation low-pass filter or modulation filterbank?,'' J. Acoust. Soc. Am. **97**, 3273 (A) (1995)].

[b)]Corresponding author. Electronic mail: torsten@medi.physik.uni-oldenburg.de

plitude modulation, using broadband noise as a carrier. Using narrow bands of noise as the masker modulation, the modulation-detection threshold function showed a peak at the masker modulation frequency. This indicates that masking is most effective when the test modulation frequency falls within the masker-modulation band. In the same vein, Bacon and Grantham (1989) found peaked masking patterns using sinusoidal masker modulation instead of a noise band. Fassel (1994) found similar masking patterns using sinusoids at high frequencies as carriers and sinusoidal masker modulation.

For spectral tone-on-tone masking, effects of frequency selectivity are well established and are associated with the existence of independent frequency channels. When translated to the modulation-frequency domain, the data of Houtgast and of Bacon and Grantham suggest the existence of modulation-frequency specific channels at a more central stage in the auditory system than the peripheral auditory filters. Yost *et al.* (1989) also suggested amplitude modulation channels to explain their data on modulation-detection interference and to account for the formation of auditory ''objects'' based on common modulation. Similarly, Martens (1982) proposed that the auditory system realizes some kind of short-term spectral analysis of the temporal waveform of the signal's envelope.

Modulation-frequency specificity has also been observed in different physiological studies of neural responses to amplitude modulated tones (Creutzfeldt *et al.*, 1980; Langner and Schreiner, 1988; Schreiner and Urbas, 1988; Langner, 1992). Langner and Schreiner (1988) stated that the auditory system contains several levels of systematic topographical organization with respect to the response characteristics that convey temporal modulation aspects of the input signal. A general reduction in the temporal activity patterns of neural elements along the auditory pathway was described as the most basic temporal organizational feature. That is, the temporal resolution of the auditory nerve (Palmer, 1982) appears to be higher than at any other processing level. The highest best modulation frequencies (BMF) found in the inferior colliculus (IC)—which is about 1000 Hz—are still comparable with the temporal resolution of auditory nerve fibers (Schreiner and Langner, 1984; Langner and Schreiner, 1988). However, the majority of units in the IC are tuned to modulation frequencies well below the upper frequency limit given by the auditory nerve. All estimates of temporal resolution in the IC were found to be higher than estimates in the auditory cortex (which are in the range of BMF=0–20 Hz in cats). Thus the auditory cortex seems to be limited in its ability to follow fast temporal changes in the input envelope. On the other hand, the cortex seems to be capable of processing slow modulations like rhythmlike envelope fluctuations (Creutzfeldt *et al.*, 1980). A further organizational level of the temporal processing is reflected by differences found in various subdivisions of auditory nuclei. For example, Langner and Schreiner (1988) found a highly systematically organized *map* of best modulation frequencies within the IC of the cat. Overall, Langner and Schreiner (1988) concluded that temporal aspects of a stimulus, such as envelope variations, represent a major organizational principle of the auditory system, that complements the well-established spectral (tonotopic) and binaural organization.

The present psycho-acoustical study further analyzes the processing of amplitude modulation in the auditory system. The goal is to gather more information about modulation-frequency selectivity and to set up corresponding simulations with an extended version of a model of the ''effective'' signal processing in the auditory system, which was initially developed to describe temporal masking effects (Dau *et al.*, 1996a, b). As already pointed out, in most classical studies of temporal processing, a broadband noise carrier has been applied to determine the TMTF. Unfortunately, the use of broadband noise carriers does not provide direct information about spectral effects in temporal processing: Broadband noise excites a wide region of the basilar membrane, leaving unanswered the question of what spectral region or regions are being used to detect the modulation. For this reason, measurements and corresponding simulations with stochastic narrow-band noises as carriers at a high center frequency were performed, as was done earlier by Fleischer (1982a, 1983). At high center frequencies, the bandwidth of the auditory filters is relatively large so that there is a larger range of modulation rates over which the sidebands resulting from the modulation are not resolved. Instead, the modulation is perceived as a temporal attribute, like fluctuations in loudness (for low modulation rates) or as roughness (for higher modulation rates). The bandwidth of the modulated signal was chosen to be smaller than the bandwidth of the stimulated peripheral filter. This implies that all spectral components are processed together and that temporal effects are dominant over spectral effects.

I. DESCRIPTION OF THE MODEL

In Dau *et al.* (1996a), a model was proposed to describe the effective signal processing in the auditory system. This model allows the prediction of masked thresholds in a variety of simultaneous and nonsimultaneous conditions (Dau *et al.*, 1996b). It combines several stages of preprocessing with a decision device that has the properties of an optimal detector. Since then, the model has also been used to predict speech perception tasks, such as automatic speech recognition and speech quality evaluation (cf. Holube and Kollmeier, 1996; Kollmeier *et al.*, 1996). Figure 1 shows the extended model that is proposed to describe experimental data on modulation perception. Instead of the implementation of the basilar-membrane model developed by Strube (1985), as used in Dau *et al.* (1996a), the gammatone filterbank model of Patterson *et al.* (1987) is used to simulate the bandpass characteristics of the basilar membrane. The gammatone filterbank has the advantages that its algorithm is much more efficient than the Strube model and that the bandwidths more closely match estimates of auditory-filter bandwidths. The signal at the output of a single filter of the gammatone filterbank is half-wave rectified and low-pass filtered at 1 kHz, as in the model described in Dau *et al.* (1996a).

The subsequent nonlinear adaptation stage is a slightly modified version (Münkner, 1993) of the adaptation stage (Püschel, 1988) implemented within the masking model of Dau *et al.* (1996a). In this modified version the amplitude of

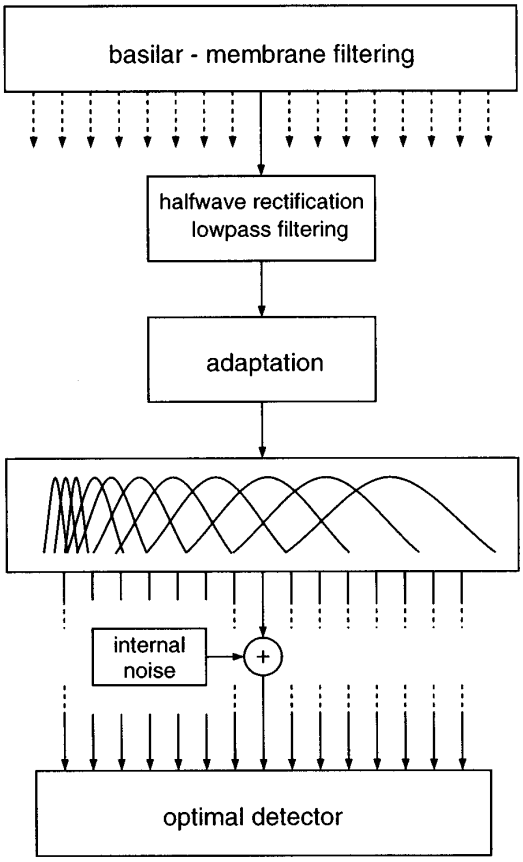

FIG. 1. Block diagram of the psycho-acoustical model for describing modulation-detection data with an optimal detector as decision device. The signals are preprocessed, subjected to adaptation, filtered by a modulation filterbank and finally added to internal noise; this processing transforms the signals into their internal representations.

the onset response was limited to a value of maximally 10 times the value of the steady state response of the stage (for details see Münkner, 1993).[1] With regard to the transformation of envelope variations of the signal, the adaptation stage transforms rapid input variations (as compared with the time constants of the low-pass filters) *linearly*. If these changes are slow enough then, because of the time constants of the model, the gain is also changed. Each element within the adaptation model combines a static compressive nonlinearity with a higher sensitivity for fast temporal variations (for details, see Dau *et al.*, 1996a).

The following stage in the model, as shown in Fig. 1, contains the most substantial changes compared to the model described in Dau *et al.* (1996a). Instead of the low-pass filter with a cutoff frequency of 8 Hz, a linear filterbank is assumed to further analyze the amplitude changes of the envelope. This stage will be called the modulation filterbank throughout this paper. A first implementation of such a modulation filterbank was presented in Fassel and Püschel (1993) and Münker and Püschel (1993). The implementation of this stage is in contrast to the signal processing within other models in the literature (e.g., Viemeister, 1979; Forrest and Green, 1987).

It is postulated within the present model that the modulation filterbank exhibits two domains with different scaling. Figure 2 shows the transfer functions of the modulation filters. In the range 0–10 Hz a constant bandwidth of 5 Hz is assumed. The lowest modulation filter represents a low-pass filter with a cutoff frequency of 2.5 Hz. From 10 Hz up to 1000 Hz a logarithmic scaling with a constant Q value of 2 is assumed.[2] The spacing in the modulation-frequency domain resembles the spacing of critical bands in the audio-frequency domain. Within the model only the (Hilbert) *envelope* of the modulation filter outputs for center frequencies above 10 Hz is further examined, introducing a nonlinearity in the processing of amplitude modulation.[3] For filters with a lower center frequency it is not reasonable to extract the Hilbert envelope from the signal, because the distinction between carrier and envelope becomes ambiguous due to the large relative bandwidth of these filters. Furthermore, the successful description of masking data by the original model version in Dau *et al.* (1996b) suggests that use is made of information about modulation phase at low modulation rates. In this model the signal envelope was analyzed by the simple 8-Hz low-pass filter and this filtering preserves all information about the modulation phase for low modulation frequencies. The present model thus tries to find a ''link'' between the description of phenomena of modulation detection and those of the more common signal detection.

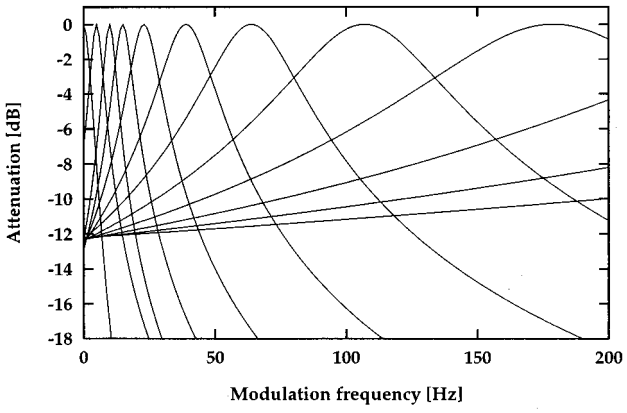

FIG. 2. Transfer functions of the modulation filters. In the range 0–10 Hz the functions have a constant bandwidth of 5 Hz. Between 10 and 1000 Hz a logarithmic scaling with a constant Q value of 2 is applied. Only the range from 0 to 200 Hz is plotted.

The output of the ''preprocessing'' stages can now be interpreted as a three-dimensional, time-varying activity pattern. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation filter output.[4] The internal noises at the outputs of the different modulation channels are assumed to be independent of each other. For stochastic input signals, the outputs of the modulation channels are not (fully) uncorrelated because of the overlap of the modulation filters. The transformed signal after the addition of noise is called the internal representation of the signal. The decision device is realized as an optimal detector in the same way as described in Dau *et al.* (1996a, b). There, the decision device of the model was first described for masking conditions using sinusoidal test signals presented in a frozen-noise masker. In each interval of a simulated 3-interval forced-choice (3IFC) adaptive paradigm, the difference between the current representation and the ''stored'' internal representation of the deterministic

masker was calculated, leading to two intervals containing only internal noise and one interval containing the nonlinearly transformed signal plus internal noise. To apply the model to random noise, a different sample of the masker was presented in each interval; the ''reference'' representation was modeled by calculating the mean internal representation of several masker samples (cf. Dau *et al.*, 1996b). This averaged reference was computed once before the adaptive procedure was started. During the adaptive procedure, the three difference representations (in a trial) were affected by the statistical properties of the external noise in addition to the internal noise. Such an algorithm is also used for the present study dealing with stochastic noise as the carrier (and reference) and a sinusoidal modulation (as the test signal). The template is generated in the present study as the normalized difference between an averaged suprathreshold internal representation of several modulated carrier samples and the averaged internal representation of the reference alone (cf. Dau *et al.*, 1996b). The decision criterion within the optimal decision stage (see Green and Swets, 1966) is given by the difference between the largest cross-correlation coefficient of the two carrier-alone representations with the template and the correlation value between the representation in the signal interval and the template. When this difference is smaller than the limit of resolution determined by the internal noise, the test modulation is not detected (for details, see Dau *et al.*, 1996b).[5]

II. METHOD

A. Procedure and subjects

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

B. Apparatus and stimuli

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

Several modulation-detection and modulation-masking experiments were performed. In most measurements narrow-band Gaussian noise centered at 5 kHz was used as the carrier. In the masking experiment a sinusoidal carrier at 5 kHz was used. The carrier level was 65 dB SPL in both cases. The specific choice of the parameters for the windowing of the stimuli will be described later in the paper when the corresponding experiments are discussed. In the experiments using a noise carrier, an independent sample of noise was presented in each interval. With one exception described below, the noise stimuli were digitally filtered before modulation by setting the magnitude of the Fourier coefficients to zero outside the desired passband.

An amplitude-modulated noise has sidebands that extend $\pm f_m$ Hz from the edges of the passband of the unmodulated noise, where f_m indicates the modulation frequency. In principle it is possible that the detection of modulation is based on spectral changes in the modulated waveform. The usability of these spectral cues depends on frequency region, owing to the relation of frequency difference limens and center frequency (Wier *et al.*, 1977; Eddins, 1993). One way to avoid these spectral cues is to apply the modulation to wide-band noise before bandpass filtering. In the present study this was done for the largest applied carrier bandwidth, 314 Hz, by setting the magnitude of the Fourier coefficients to zero outside the desired passband. This is the same procedure that was applied by Eddins (1993). Thus in this case, the bandwidth of the stimuli is the same regardless of the presence or absence of modulation. By filtering after amplitude modulation, the sidebands introduced by modulation are effectively reduced. The filtering after modulation causes a partially filling in the valleys of the temporal waveform (e.g., Eddins, 1993). However, this technique ensures that spectral cues were not available and the task was purely temporal in nature. In contrast, for the carrier bandwidths of 3 and 31 Hz, no filtering after modulation was applied.

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

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

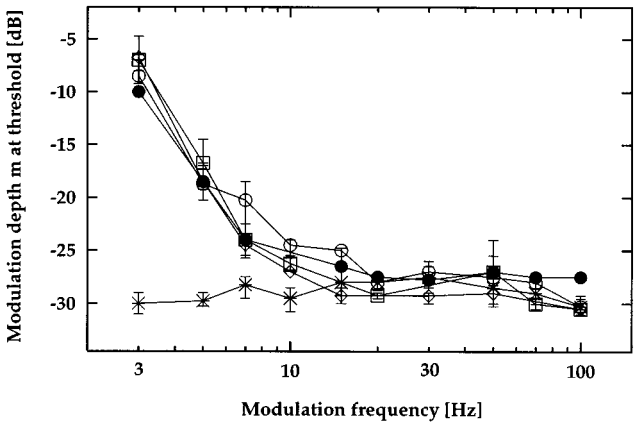

FIG. 3. Modulation-detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 3-Hz-wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: JV (□); AS (◇); TD (○); optimal detector (●).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

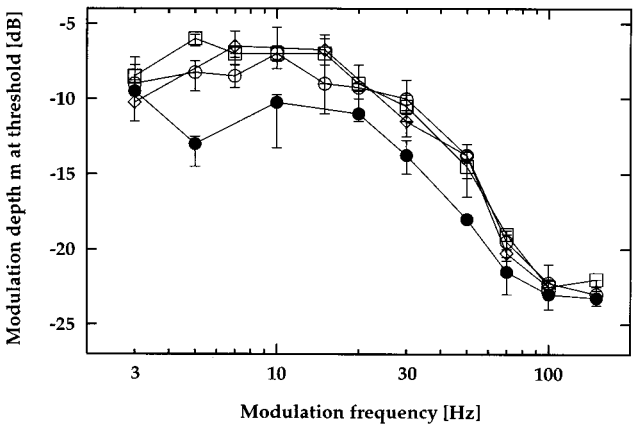

FIG. 4. Modulation-detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 31-Hz-wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: AS (◇); TD (○); JV (□); optimal detector (●).

III. RESULTS

A. Amplitude-modulation thresholds of narrow-band noise as a function of the carrier bandwidth

In this experiment, TMTFs were measured and simulated for narrow-band noise carriers of bandwidths 3, 31, and 314 Hz, centered at 5 kHz in each condition (cf. Fleischer, 1982a, 1983). Fleischer's experiments were replicated in this study and compared with corresponding simulations carried out with the present model. In contrast to Fleischer, an adaptive threshold procedure was used and the carrier level was somewhat lower (65 dB SPL). For the three noise bandwidths, the corresponding spectrum levels were about 60, 50, and 40 dB SPL. The carrier and the applied sinusoidal modulation had a duration of 1 s. Both were windowed with 200-ms cosine-squared ramps. Figure 3 shows the present experimental results for amplitude modulation detection employing a carrier bandwidth of 3 Hz at a center frequency of 5 kHz. The figure shows the data of three subjects (open symbols) together with the model predictions (closed symbols). For comparison, data obtained by one subject using a sinusoidal carrier at 5 kHz are shown as asterisks. For the sinusoidal carrier, the same stimulus parameters were used as for the noise carrier conditions. The ordinate indicates modulation depth at threshold, and the abscissa represents the modulation frequency. A comparatively high detection threshold is observed at a modulation rate of 3 Hz. This is due to the *inherent* statistical fluctuations of the narrow-band 3-Hz-wide carrier. These inherent fluctuations of the carrier envelope mask the additional periodic 3-Hz test modulation. With increasing modulation frequency, thresholds decrease and converge with those obtained using a sinusoidal carrier at a modulation frequency of 20 Hz. The threshold remains flat up to a modulation frequency of 100 Hz. This finding indicates that the auditory system does not seem too slow or sluggish to follow fast fluctuations in this range. There is very good agreement between the measurements and simulations. The flat threshold function up to 100 Hz contrasts with the conclusions derived from modulation detection data for broadband noise carriers (e.g., Viemeister, 1979) and also contrasts with data in Zwicker (1952) obtained with a sinusoidal carrier. Zwicker found an increase in threshold between 4 and 64 Hz of about 9 dB at a carrier frequency of 4 kHz. However, the present data are in good agreement with more recent data by Fleischer (1982a, 1983), Fassel (1994), Fassel and Kohlrausch (1995), Dau (1996), and Fassel *et al.* (1997), who measured TMTFs with sinusoidal carriers at 5 and 10 kHz.

Flat thresholds up to a modulation frequency of 128 Hz were also observed by Strickland and Viemeister (1997) in an experiment where subjects had to discriminate between AM and quasi-frequency modulation (QFM) applied to a sinusoidal carrier of 4 kHz. Based on additional data on QFM detection, these authors argued that the flatness in their TMTF between 64 and 128 Hz may have been caused by the increasing role of spectral cues and thus did not reflect true temporal processing. Since the assumptions about available cues are of relevance for the interpretation of our data, we will return to the arguments put forward by Strickland and Viemeister in the discussion (Sec. V) of the present paper.

Figure 4 shows thresholds using a narrow-band carrier with a bandwidth of 31 Hz. Again, the modulation depth, m, at threshold was measured and simulated as a function of the test-modulation frequency. The open symbols represent the measured data of three subjects and the filled symbols indicate the simulated thresholds. The threshold at a very low modulation rate (3 Hz) is several dB lower than in the case of the 3-Hz-wide carrier. This decrease is due to the decreasing spectral energy density in the modulation spectrum with increasing bandwidth of the carrier (see the Appendix). In terms of the model, less ''noise energy'' falls into the low-frequency modulation filter which is tuned to the test-modulation frequency. For modulation frequencies larger than half the bandwidth of the noise ($f_{\mathrm{mod}} > \Delta f/2$) thresholds begin to decrease, both in the measurements and in the simulations, so that a high-pass characteristic in the threshold function becomes apparent. However, thresholds decrease more slowly with increasing modulation frequency than the

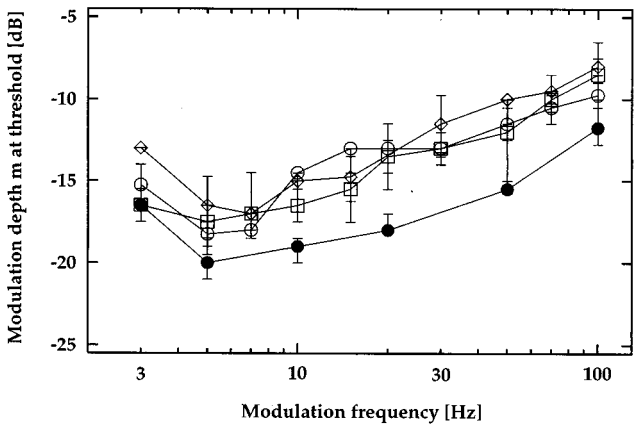

FIG. 5. Modulation-detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 314-Hz-wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: JV (□); TD (○); AS (◇); optimal detector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

spectrum of the inherent envelope fluctuations itself (e.g., Lawson and Uhlenbeck, 1950). The idealized modulation spectrum of a rectangular shaped bandpass noise has a triangular shape which stretches from 0 to Δf Hz (see the Appendix and Lawson and Uhlenbeck, 1950). If the auditory system would be sharply tuned in frequency selectivity for modulation, thresholds would decrease with increasing modulation frequency more or less in parallel with the spectral shape of the modulation spectrum of the carrier (assuming a constant signal-to-noise ratio at the output). Apparently, this is not the case. Hence, even at high modulation rates of 100 and 150 Hz, thresholds have not yet converged with those for the 3-Hz-wide carrier nor with those for the sinusoidal carrier, but are about 5 dB higher. This implies that the relatively slow inherent fluctuations of the 31-Hz-wide carrier make it difficult to detect the higher-frequency test modulation. This phenomenon was also observed by Fleischer (1982a) who referred to it as ''cross-talk'' of the inherent fluctuations of the noise on the added modulation. This effect decreases with increasing rate of the test modulation.

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

Figure 5 shows results for the carrier bandwidth of 314 Hz. Thresholds are higher for a modulation rate of 3 Hz than for a rate of 5 Hz. This is probably caused by the use of a gated carrier. Such an effect has been observed in several psycho-acoustical studies (e.g., Viemeister, 1979; Sheft and Yost, 1990). Based on their results it can be assumed that the threshold at 3 Hz would decrease if a continuous carrier had been used instead of a gated one. For modulation frequencies above 7 Hz, thresholds increase by about 3 dB per doubling of the modulation frequency. This threshold pattern agrees well with comparable experimental data of Eddins (1993) for a carrier bandwidth of 400 Hz. The form of the TMTF is similar to the pattern found in ''classical'' measurements of the TMTF using a broadband noise as a carrier, but it has a much lower cutoff frequency (Viemeister, 1979). Overall, the threshold curve is very different from those obtained with smaller carrier bandwidths since the detectability of the test modulation decreases with increasing modulation frequency. Consistent with the data, the simulations also show increasing thresholds with increasing modulation frequency.

For a carrier bandwidth of 314 Hz, the spectrum of the intrinsic fluctuations is relatively flat over the whole range of the test-modulation frequencies. The additional test component falls in the passband of mainly one modulation filter. Assuming a constant decision criterion at threshold, the logarithmic scaling of the modulation filters with center frequencies above 10 Hz leads to an approximately 3-dB increase of modulation depth, m, at threshold per doubling of modulation frequency. In other words, to get the same signal-to-noise ratio at threshold, a greater modulation depth is required with increasing modulation frequency. Thus the apparent modulation low-pass behavior in the model data in Fig. 5 is not explained by assuming a general low-pass characteristic in the auditory system, but is caused by the constant relative width (or logarithmic scaling) of the modulation filters.

Figure 6 gives an illustration of how the signals are internally represented in the model. It shows how the template is derived from the internal representation of suprathreshold test modulation and that of the unmodulated carrier alone. The upper panel shows the three-dimensional internal representation of a 3-Hz-wide carrier alone (centered at 5 kHz). It represents the internal activity as a function of time and center frequency of the modulation filters. The ordinate is scaled in model units (MU). The modulation center frequencies range from 0 to 1000 Hz. Since the total energy within the modulation spectrum of the signal is concentrated at very low modulation rates, only the lowest modulation filters are excited by the input signal. This is indicated by the hatched lines in the figure. At the beginning of the carrier, all modulation filters show a short period of high excitation. This response reflects the step response of the filters to the envelope onset. The middle panel of Fig. 6 shows the internal representation of the carrier, this time sinusoidally modulated with a test-modulation rate of 20 Hz at a highly detectable modulation depth. The test modulation mainly activates the modulation filter tuned to 20 Hz but also stimulates adjacent modulation filters, because of the relatively low modulation-frequency selectivity assumed in the model. Again, the inherent fluctuations of the carrier itself primarily activate the region at low modulation frequencies. However, because of the large spectral separation between the test modulation and the inherent fluctuations of the carrier, there is no interaction between the two components; that is, no competing ''noise'' energy leaks into the transfer range of the test-modulation filter. The lower panel in Fig. 6 gives the

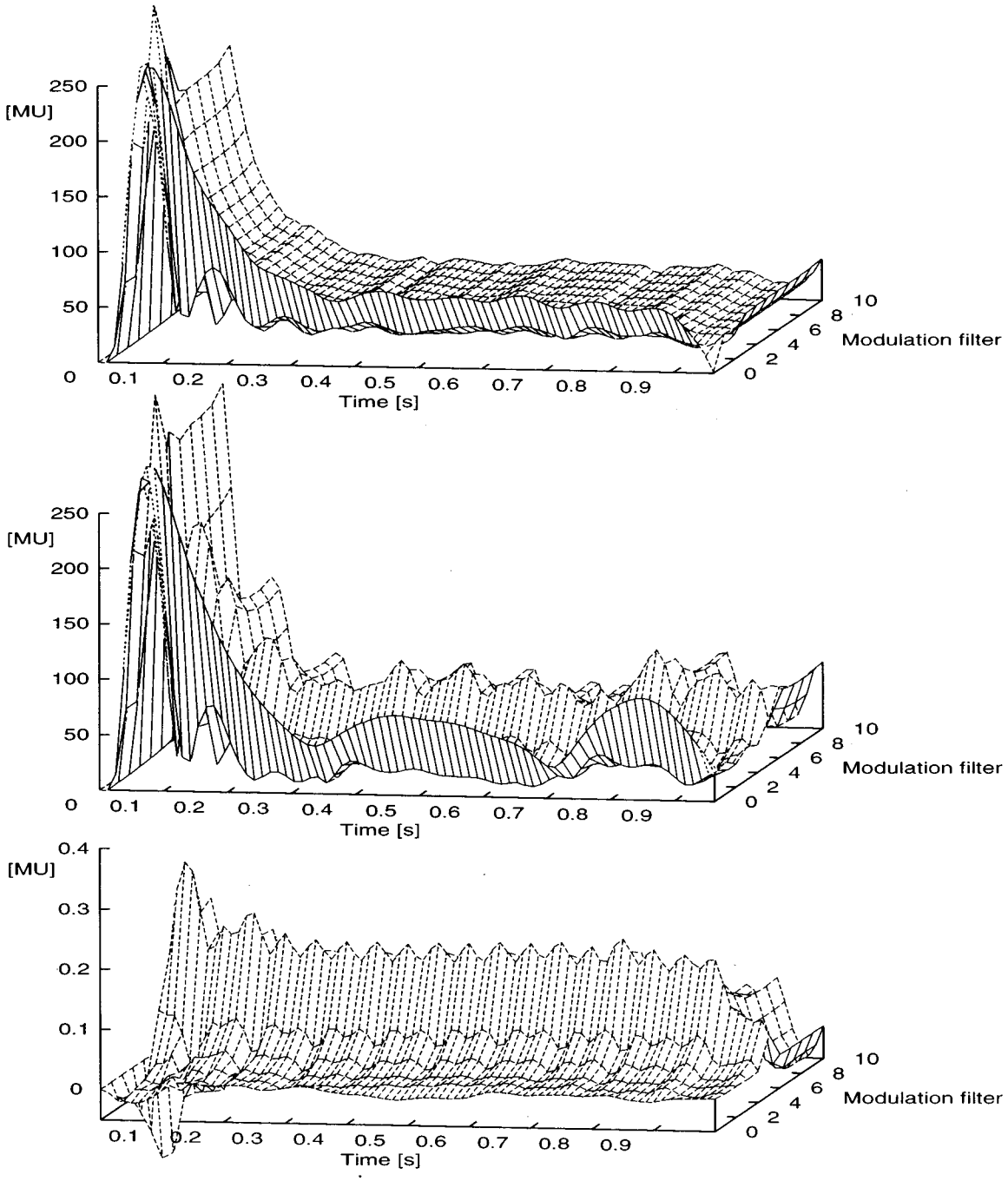

FIG. 6. Generation of the template representation (at the bottom) of a 20-Hz test modulation which was impressed on a 3-Hz-wide running noise carrier centered at 5 kHz. The template is the normalized difference between the mean representation of the carrier plus the suprathreshold modulation (in the middle) and the mean representation of the carrier alone (at the top). The ordinate is scaled in model units (MU).

template which is derived by subtracting the upper panel from the middle one and normalizing the result. As a consequence of the marked separation in modulation frequency, the internal representation of the template contains a representation of the temporal course of the test modulation without interference from the carrier modulation.

B. Amplitude modulation thresholds of third-octave-wide noisebands as a function of the center frequency

In the previous section it was observed that the detection threshold for amplitude modulation depends on the spectral density of the inherent fluctuations of the carrier, when the total energy of the modulated signal is constant. This is examined further in the following experiment using a third-octave-wave noiseband as the carrier. The detection threshold for 25-Hz modulation was measured and simulated as a function of the center frequency of the band. Stimulus parameters were the same as in the experiments of the previous section. In the model, only the output of the peripheral filter centered on the bandpass noise was analyzed. It was further assumed that the scaling of the modulation filters does not change with the peripheral frequency region. Figure 7 shows the modulation depth at threshold as a function of the center

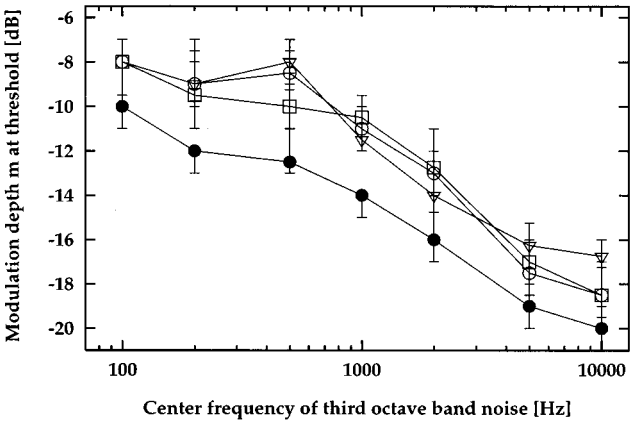

FIG. 7. Modulation-detection thresholds for a 25-Hz modulation as a function of the center frequency of a third-octave-wide noise carrier. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: Data from Fleischer (1981) (□); TD (○); JV (▽), optimal detector (●).

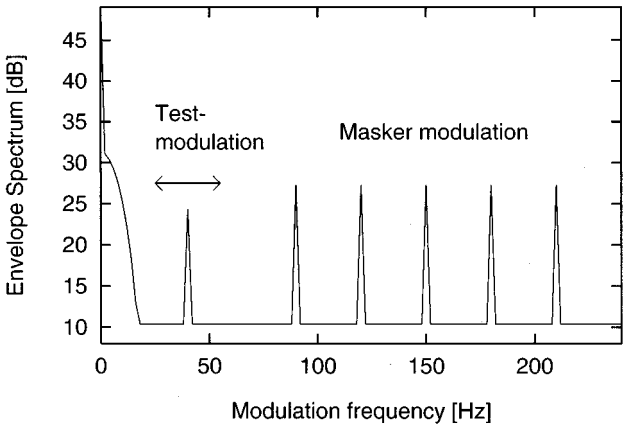

FIG. 8. Logarithmic spectrum of the Hilbert envelope of the stimuli presented in the signal interval of the modulation masking experiment. The signal interval contains the five components of the masking tone complex and the signal component. The subject's task was to detect the signal component.

frequency of the third-octave-wide noiseband. Modulation thresholds decrease with increasing center frequency. The increasing absolute bandwidth results in a decreasing density of inherent low-frequency envelope fluctuations, if the total energy of the modulated stimulus is kept constant. As a consequence, less energy from the random envelope fluctuations of the carrier falls within the passband of the modulation filter tuned to the test-modulation frequency. This leads to decreasing thresholds with increasing center frequency in the model. Apart from the systematic 2- to 3-dB difference in the absolute sensitivity, there is good agreement between the simulated and the measured data.

C. Modulation masking: a harmonic tone-complex masker

In a further experiment concerning modulation-frequency selectivity, a masking paradigm for investigating frequency selectivity in the audio-frequency domain was adopted. It served as a test for spectral analysis in the modulation domain, as opposed to a periodicity analysis. The carrier was a 5-kHz sinusoid. A narrow-band tone complex was used as masker modulation. This complex consisted of the third through seventh components of a harmonic tone complex with a fundamental frequency of 30 Hz, with frequencies of 90, 120, 150, 180, and 210 Hz. The amplitude of each component was 0.16, a value sufficiently low to avoid overmodulation when the test modulation was combined with the tone-complex modulation. In each interval, the starting phase of each spectral component was randomly chosen from a uniform distribution in the range 0–360°. With this choice, the modulating tone complex had a noiselike, but periodic, waveform. A sinusoidal test modulation was imposed on the same carrier. The test modulation was chosen from the range 20–120 Hz. Thus the bandwidth of the modulated signal remained within the bandwidth of the auditory filter centered at 5 kHz. Figure 8 shows schematically the spectral distribution of the masker and test components in the modulation spectrum. The modulated stimuli were presented at a level of 65 dB SPL, and had a duration of 400 ms. Test and masker modulation were present for the whole duration of the carrier and were gated with 20-ms cosine-squared ramps.

The amount of modulation masking as a function of the test-modulation frequency is shown in Fig. 9. The unmasked modulation thresholds, i.e., the thresholds for sinusoidal test modulation without any interfering masker modulation, were used as a reference to evaluate the effect of the modulated masker. These reference thresholds were similar across subjects and were similar to those described in the first experiment of this paper (see Fig. 3), remaining more or less flat up to a modulation frequency of 120 Hz. The ''masking pattern'' was derived by subtracting the unmasked threshold from the masked threshold at each test-modulation frequency. As can be seen from Fig. 9, the amount of masking increases with increasing test-modulation frequency. The difference between the highest and the lowest threshold was more than 10 dB. Note that there is no peak at 30 Hz, the ''missing fundamental.'' This indicates that the masking effect is not determined by the period of the masker modulation. Also, no pronounced peak in threshold is observed for

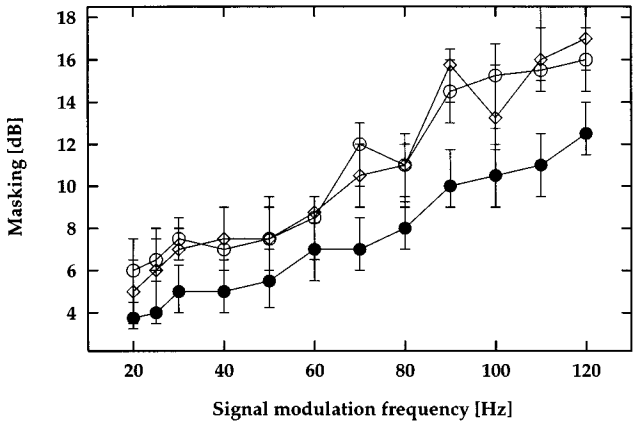

FIG. 9. Amount of modulation masking as a function of the modulation frequency. Carrier: 5-kHz sinusoid, modulation masker: 3rd–7th components of a harmonic tone complex with fundamental $f_0 = 30$ Hz. Level: 65 dB SPL. Subjects: TD (○); JV (◇); optimal detector (●).

modulation frequencies of 90 and 120 Hz, corresponding to the lowest harmonic components of the masker complex. This indicates that no sharp tuning in modulation frequencies occurs at these comparatively high modulation frequencies. An increasing masking effect with increasing test modulation frequency is also seen in the simulations (filled symbols in Fig. 9). If masking effects in the modulation-frequency domain were determined by the periodicity of the stimuli one would expect an increased threshold at the fundamental frequency of the tone complex and at higher harmonics. If, however, the masked threshold of the test signal were mainly determined by the auditory system's frequency selectivity for modulation, one could conclude that the system performs a modulation analysis which is comparable and analogous to the ''critical-band'' filtering on the basilar membrane. The experimental data clearly suggest the latter case. The simulations show good agreement with the experimental data. However, there is a systematic difference of 2–4 dB between the measured and simulated masking patterns. The masked threshold is directly related to the amount of masker energy falling within the passband of the modulation filter tuned to the actual test modulation. For the lowest test modulation rate (20 Hz) there is only a very small masking effect in the model since the modulation filters in the low modulation-frequency region are assumed to be relatively sharply tuned (see Fig. 2). With increasing test-modulation frequency, more and more components of the masker contribute to masking. Also, in the simulations the difference between the highest and the lowest masked threshold amounts to nearly 10 dB. These results further support the notion of modulation-frequency selectivity, although this selectivity seems to be relatively broadly tuned.

IV. COMPARISON WITH PREDICTIONS OF VIEMEISTER'S MODEL FOR MODULATION DETECTION

The modulation filterbank concept differs considerably in its structure from the ''classical'' modulation low-pass filter approach (e.g., Viemeister, 1979). In this section, predictions of the modulation low-pass filter approach are investigated and compared with the performance of the modulation filterbank model.

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

Figure 10 shows simulated TMTFs for noise carriers of 3-, 31-, 314-, 2000-, and 6000-Hz bandwidth on the basis of Viemeister's model. The narrow-band stimuli were the same as in Figs. 3–5. All curves show a low-pass characteristic

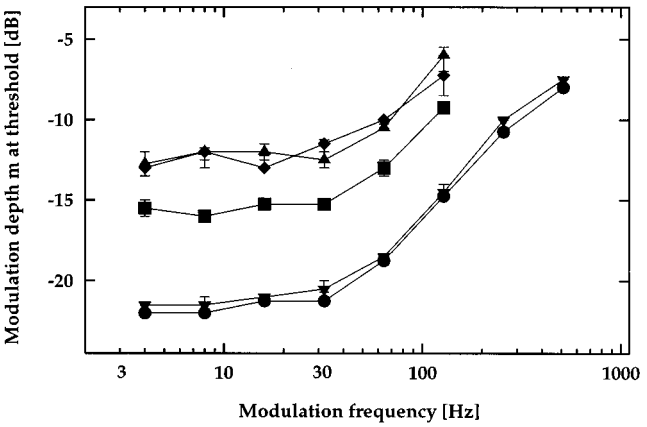

FIG. 10. Simulations on the basis of Viemeister's model. Predicted modulation detection thresholds are shown for five different bandwidths of the noise carrier. Center frequency of the carrier: 5 kHz. Carrier bandwidth: ♦: 3 Hz; ▲: 31 Hz; ■: 314 Hz; ▼: 2000 Hz; ●: 6000 Hz.

with a similar cutoff frequency. This characteristic reflects the influence of the low-pass filter stage. With decreasing carrier bandwidth, the simulated TMTFs shift toward higher thresholds. For the output of Viemeister's model, this increase will be seen for carrier bandwidths that are less than the bandwidth of the predetection filter and greater than the cutoff frequency of the low-pass filter. For these conditions, the ac-coupled rms value of the unmodulated noise carrier at the output of the modulation low-pass stage increases with decreasing carrier bandwidth.

At very low modulation frequencies, the increase in threshold agrees qualitatively with the experimental data. At higher modulation frequencies, however, the model predicts a totally different threshold pattern than that observed experimentally.

While the pattern of the experimental data varies systematically with increasing carrier bandwidth, the model always predicts a low-pass characteristic in the threshold function independent of the carrier bandwidth. Note that the model proposed here provides a better description of the experimental data (cf. see Sec. III A, Figs. 3–5).

Figure 11 shows model predictions of amplitude modu-

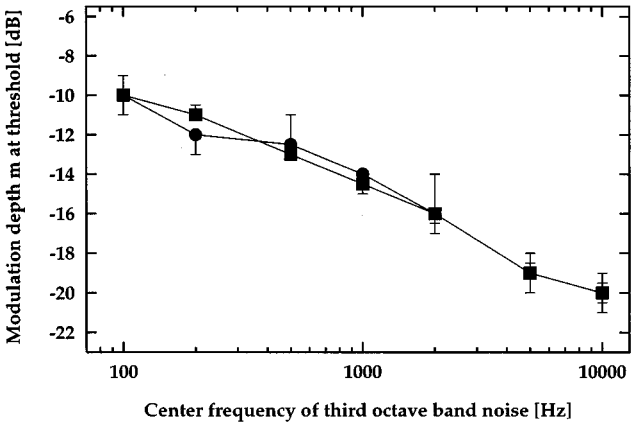

FIG. 11. Simulated modulation detection thresholds for 25-Hz amplitude modulation as a function of the center frequency of the third-octave-wide noise carrier. Viemeister model: (■); modulation filterbank model (●).

lation detection thresholds of third-octave-noise bands as a function of the center frequency. The stimuli were the same as in Sec. III B. The filled squares represent thresholds on the basis of the Viemeister model. Thresholds decrease monotonically with increasing center frequency. This is again caused by the decreasing ac-coupled rms value of the unmodulated noise with increasing center frequency (and increasing linear bandwidth) at the output of the modulation low-pass stage, in the same way as described above. The filled circles in the figure show the simulated data obtained with the modulation filterbank model (replotted from Fig. 7). There is virtually no difference between the predictions of the two models. The predicted threshold functions account well for the data (see Fig. 7).

Finally, model predictions were calculated for the modulation-masking experiment with the harmonic tone complex as the masker. The stimuli were the same as in Sec. III C. The model of Viemeister predicts about 5-dB masking for *all* test-modulation frequencies. This frequency-independent masking is caused by the specific model structure: Because there is only one modulation-frequency specific channel (the output of the low-pass filter) the effect of the masker modulation is the same for all test modulation frequencies at least in combination with the decision algorithm proposed by Viemeister (1979). Hence the experimentally observed increase in modulation masking with decreasing spectral distance between modulation masker and test modulation cannot be described properly with Viemeister's model.

While this paper was being written, a recent article by Strickland and Viemeister (1996) showed that by replacing the rms criterion with a max/min statistic, a single-channel envelope detector can capture some aspects of modulation masking data, so that the effect of masker modulation is not the same at all test modulation rates. However, their model predictions showed a much too sharp ''tuning'' to modulation frequency compared to the relatively broadly tuned masking patterns found in the data (Houtgast, 1989; Bacon and Grantham, 1989). In order to better understand the properties of such statistics, we repeated the simulations shown in Fig. 10 using a max/min decision device. Basically, the change in detector criterion from rms to max/min does not change the main aspects of the previously discussed curves: Independent of carrier bandwidth, all TMTFs have the same shape and increase with increasing modulation frequency. For a reduction of the carrier bandwidth between 2000 and about 30 Hz, the TMTFs are shifted toward higher threshold values. In addition, as already mentioned by Forrest and Green (1987), the max/min statistic is less stable than the rms statistic.

In summary, the analysis of various models proposed in the literature and the simulations from the present study provide a strong indication for a modulation-frequency specific analysis in the auditory system. The modulation filterbank model, which is able to reproduce at least the trend in the data, is one possible realization for this analysis.

V. DISCUSSION

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

With regard to the experiments performed and the structure of the model that is inferred from these data, the following points should be discussed: (a) the assumption that the use of spectral cues, effects of peripheral filtering, and off-frequency listening can be neglected for the conditions tested in this paper; (b) the concept of a modulation filterbank as opposed to a modulation low-pass filter in each critical band; and (c) the envelope statistics of the different noise maskers employed and their influence on the thresholds obtained here.

A. Role of spectral cues, peripheral filtering, and off-frequency listening

The experiments in this study have been designed so as to minimize effects of spectral cues and peripheral filtering. The carrier frequency was very high and therefore the bandwidth of the modulated signals was always smaller than the bandwidth of the stimulated peripheral filter. We also argued that at a carrier frequency of 5 kHz, temporal cues are dominant over spectral cues in modulation detection for modulation frequencies up to at least 100 Hz, a view we find supported by measurements with sinusoidal carriers (see Sec. III A, and Fleischer, 1982a, 1983; Fassel and Kohlrausch, 1995, 1996; Dau, 1996; Fassel *et al.*, 1997).

A somewhat different view about the flatness of TMTFs for tonal carriers at 4 kHz was put forward in a recent paper by Strickland and Viemeister (1997). Based on data for discrimination between AM and QFM, and on data for detecting QFM, they argued that the thresholds for AM vs QFM at a modulation frequency of 128 Hz were not just caused by temporal cues (as we assume), but that other, probably spectral cues, were also involved. It is further implied that this may also be the reason for the flatness in tonal TMTFs at 3 and 5 kHz in Fassel and Kohlrausch (1996) and Dau (1996), which would undermine one of the assumptions used in our interpretation. In the following we argue why we find this implication not convincing.

According to Strickland and Viemeister, the sensitivity to temporal cues alone would lead to an increase in thresholds for AM vs QFM discrimination above 64-Hz modulation frequency. Only due to the availability of additional cues, thresholds appear to be flat up to 128 Hz. The usability of these additional cues is derived from experiments measuring QFM detection. At a 4-kHz carrier frequency and 128-Hz modulation frequency, QFM detection thresholds are about 8 dB higher than the thresholds for discriminating AM from QFM. These differences in level make it, in our view, very difficult to see room for a reasonable contribution of nontem-

poral cues, unless it was shown that the psychometric function for QFM detection was sufficiently shallow.

If we compare AM detection thresholds at a wider range of carrier frequencies, the interpretation by Strickland and Viemeister leads to the following view. Data by Zwicker (1952) show that QFM detection thresholds as a function of modulation frequency decrease earlier for lower than for higher carrier frequencies and that the shift in shape is about 1 oct in modulation frequency per 1 oct in carrier frequency (cf. Figs. 10 and 11 in Zwicker, 1952). If we make the parsimonous assumption that the contribution of nontemporal cues to AM vs QFM discrimination follows the same basic rules across carrier frequencies, the modulation frequency at which nontemporal cues start to contribute should be lower than 128 Hz for carrier frequencies below 5 kHz and higher than 128 Hz for carriers above 5 kHz. Our own results on AM detection for sinusoidal carriers do not show such behavior (Fassel and Kohlrausch, 1995; Dau, 1996; Fassel *et al.*, 1997). For all carrier frequencies between 3 and 10 kHz, AM detection thresholds remain flat up to the same modulation frequency of about 100 to 130 Hz.

Another problem in interpreting AM detection thresholds is that observers may increase the relative modulation depth in the AM stimulus by positioning their ''internal observation filter'' away from the carrier frequency in such a way as to better equate the amplitude of the carrier and one of the sidebands (e.g., Goldstein, 1967). The increase in relative modulation depth resulting from listening off frequency would improve performance. We think that for 5-kHz carriers and modulation frequencies in the range 0–100 Hz, it is unlikely that this type of off-frequency listening would be advantageous, and that subjects most likely monitor the internal filter at the carrier frequency.

First of all, the (relative) frequency difference between carrier and one of the sidebands is no larger than 2%, which corresponds to about 0.15 ERB or 0.1 Bark. In order to increase the modulation depth m at the output of an off-frequency filter by 2 dB, this filter would need to have a slope of 6 dB per 100 Hz (assuming a constant slope over the spectrum of the AM stimulus). This corresponds at 5 kHz to slopes of about 33 dB/ERB or 56 dB/Bark, values clearly higher than typical estimates of auditory filter slopes.

In addition, if this effect contributes to AM detection, it should be even stronger in single-sideband detection. If only one sideband and the carrier are available, optimal placing of a filter away from the carrier will increase the degree of modulation even more than is possible for modulation detection, where the (relative) increase of one sideband is accompanied by a decrease of the other sideband. Both for 5-kHz (Dau, 1996) and 10-kHz carriers (Fassel *et al.*, 1997), we found basically flat thresholds for detecting the lower or the upper sideband up to about 100 Hz. Even more important is the observation that sideband detection thresholds for larger spectral distances first increased, before they finally decreased. We take this as an indication that monitoring an off-frequency filter cannot significantly influence TMTFs at 5 kHz for modulation frequencies up to at least 100 Hz.

Alternatively, off-frequency listening could influence modulation detection for narrow-band carriers by the increased internal modulation depth in the region of upward spread of excitation. Such a mechanism has been proposed in the past as one of the sources for the level dependence of AM detection thresholds for sinusoidal carriers (e.g., Zwicker, 1956; Maiwald, 1967b). According to Strickland and Viemeister (1997), it also affects thresholds for low modulation frequencies in the case of bandlimited noise carriers. This conclusion was based on the fact that by adding unmodulated notched noise designed to mask the region of upward spread of a bandlimited noise carrier, modulation-detection thresholds increased by up to 7 dB. Based on this result, it was argued that measuring modulation thresholds without a notched noise would not reveal true temporal processing within the auditory filter centered on the carrier.

Interestingly, the usability of nonlinear upward spread in modulation detection for noise carriers has been addressed theoretically and by model simulations earlier by Maiwald (1967b) and we will recall the relevant points here. In the region of upward spread, both the inherent fluctuations of the noise carrier *and* the applied AM will be enhanced in a similar way. As long as the intrinsic fluctuations of the carrier are the limiting factor for modulation detection, the upward spread region does not allow a better detectability than the on-frequency region. This contrasts to the situation for sinusoidal carriers, where only the applied modulation, but not the limiting (internal) noise are enhanced in the region of upward spread.

Second, if nonlinear upward spread indeed plays such a significant role as stated by Strickland and Viemeister (1997), modulation thresholds for bandlimited noise carriers should strongly increase with decreasing carrier level, since the availability of nonlinear upward spread is strongly reduced at low and medium carrier levels. Data by Maiwald (1967b) for a 127-Hz-wide noise carrier at 1 kHz show that detection thresholds for 4-Hz modulation vary by no more than 2 dB for a level variation of 60 dB. In contrast, the same level variation for a sinusoidal carrier reveals a 15-dB effect. This suggests that the results of Strickland and Viemeister were not primarily due to the masking of the upward spread of excitation.

Following from these considerations we conclude that for the conditions investigated in the present study, monitoring off-frequency filters does not contribute significantly to modulation detection and we can indeed attribute the thresholds to being based on temporal, rather than on spectral cues.

Of course, in modulation-detection conditions with carrier bandwidths larger than a critical band, the influence of peripheral filtering on the processing of modulation frequencies can no longer be neglected. For such conditions, an extension of the ''single-channel'' model is required that allows integration of signal information across frequency. Such an extension of the single-channel model to a multi-channel model, that is able to simulate effects of spectral integration in amplitude-modulation detection and masking, is described in the accompanying paper (Dau *et al.*, 1997).

B. Modulation filterbank versus modulation low-pass filter

The modulation filterbank concept proposed here is different both from the previous version of the current model (cf. Dau *et al.*, 1996a, b) and from models proposed in the literature (e.g., Viemeister, 1979). Both of these models employed some kind of a modulation low-pass filter. The previous version of the current model was developed to describe simultaneous and forward masking data and included a low-pass filter with a cutoff frequency of 8 Hz. Such low-pass filtering, however, would fail to describe experiments concerning modulation masking and would also fail to describe basic experiments concerning modulation detection with narrow-band carriers at high center frequencies (cf. Sec. III A). The present model allows the prediction of modulation data and, at the same time, preserves the capabilities of the earlier model for describing simultaneous and nonsimultaneous masking data. This is because the linear modulation low-pass filtering (with a cutoff frequency at very slow modulations) is retained in the current model and is combined with the analysis of faster modulations by a modulation filterbank. The idea behind the modulation low-pass filter approach described in the literature (cf. Viemeister, 1979) is that a ''minimum integration time'' is typically derived from the cutoff frequency of the low-pass characteristic in the threshold function as a parameter that describes the auditory system's temporal resolution (for a review, see Viemeister and Plack, 1993). Such a model is capable of predicting a variety of different experiments, for example, the TMTF for broadband noise carriers (cf. Fig. 5) and modulation thresholds in third-octave band noise at different center frequencies (cf. Fig. 11). However, a model employing only a low-pass filter fails to describe the modulation detection data for a narrow-band noise carrier (cf. Fleischer, 1982a, 1983). Furthermore, such a model fails to describe the masking data using the tone-complex modulation masker. Hence, the model proposed here considerably expands the class of experiments that can be modeled correctly while still maintaining some of the properties and predictions of the model proposed by Viemeister (1979).

C. Intrinsic fluctuations of the noise carrier

In the past, only a few studies have attempted to involve the *inherent* statistical properties of the noise carriers in explaining and modeling TMTFs (e.g., Zwicker, 1953; Maiwald, 1967a, b; Fleischer, 1981, 1982b). For example, Fleischer (1981, 1982a, b) investigated TMTFs using narrow-band noise as the carrier. He developed a model for describing the interaction between inherent fluctuations within a noise carrier and the detectability of added modulation. The ''modulation spectrum'' was weighted by a certain factor which essentially represented a low-pass characteristic. For modulation frequencies lower than half the bandwidth of the noise carrier, this model yields good agreement with experimental results for modulation detection and modulation difference limens (Fleischer, 1981, 1982a, 1982b). For modulation frequencies larger than half the bandwidth of the noise carrier, this model would always predict a low-pass characteristic in the threshold function without regard to the applied carrier bandwidth—in the same way as shown in Sec. IV for the Viemeister model. Therefore, to account for the data, Fleischer extended the model by assuming ''cross-talk'' between the inherent fluctuations of the noise and the added modulation. He postulated a decay at a rate of 16 dB per decade of the modulation frequency to account for the high-pass characteristic in the data. In order to find a description for the inherent modulation of the noise, Fleischer (1981) regarded narrow-band noise with a bandwidth Δf as a pure tone which was amplitude modulated by a continuum of equal-amplitude modulation frequencies between zero and half the bandwidth of the noise. But this assumption is not correct. It would imply that the modulation spectrum of noise has a flat rather than a triangular shape, as shown by Lawson and Uhlenbeck (1950) (see the Appendix).

Even though the exact shape of the modulation spectrum assumed by Fleischer (1981) was not correct and the subsequent explanation of the data was based on a different concept than the one described here, Fleischer's concept of cross-talk between inherent envelope fluctuations of the carrier noise and the test modulation is compatible with the bandpass analysis proposed here. Within the modulation filterbank model, the low-pass characteristic of the threshold function for conditions with $f_{\mathrm{mod}} < \Delta f/2$ does not result from a specific weighing function used to model the ''sluggishness'' of the auditory system. Instead, it is a consequence of the intrinsic envelope fluctuations of noise bands and their spectral distribution on the one hand, and of the logarithmic scaling of the postulated modulation filters on the other hand. A critical test for this interpretation would be to obtain TMTFs for noise carriers with an envelope spectrum different from that of Gaussian noise, for example, low-noise noise (Hartmann and Pumplin, 1988; Kohlrausch *et al.*, 1997).

The current model can also account for the data using very narrow-band stimuli as the carrier, describing a high-pass or bandpass characteristic in the threshold function.

VI. CONCLUSIONS

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

(2) The model of the effective signal processing in the auditory system proposed here is capable of quantitatively modeling most aspects of the experiments described. It employs a modulation filterbank for envelope analysis that exhibits a constant absolute bandwidth for low frequencies and a constant relative bandwidth for modulation frequencies above 10 Hz. Within the context of this model, the low-pass characteristic of the broadband TMTF is due to the inherent fluctuations of the carrier and constant relative bandwidth of the modulation filters, and not to a low-pass characteristic within the auditory system *per se*.

(3) While the predictions of the model proposed here agree with some predictions of the modulation low-pass model by Viemeister (1979) and an earlier version of the

current model (Dau *et al.*, 1996a, b), the model also holds for experiments such as modulation detection for narrow-band noise carriers, where the modulation low-pass approach clearly fails.

ACKNOWLEDGMENTS

We would like to thank all our colleagues of the Graduiertenkolleg Psychoakustik at the University of Oldenburg for fruitful discussions on the content of this paper. We also thank Brian Moore, Jesko Verhey, Andrew Oxenham, Stefan Münkner, and Ralf Fassel for their comments and suggestions concerning this study and for their critical reading of earlier versions of this paper. Two anonymous reviewers also provided very constructive criticism.

APPENDIX: ENVELOPE SPECTRA OF GAUSSIAN NOISES

The time-averaged power of the envelope is twice the average power of the waveform. Hence, it is independent of the noise bandwidth, as long as the total waveform power is fixed (cf. Hartmann and Pumplin, 1988). Therefore, for example, two Gaussian waveforms with the same power but with different bandwidths, have the same envelope power.

An interesting question is related to the *spectral* distribution of the envelope power. Lawson and Uhlenbeck (1950) calculated the spectrum of the envelope via Fourier transform of the autocorrelation function of the envelope. Assuming a rectangular shape of the power spectrum of the noise, they showed that the modulation spectrum $N=N(f_{\mathrm{mod}})$, i.e., the power spectrum of the (linear) envelope of the noise, is given approximately by the formula:

$$N_{\Delta f,\rho}(f_{\mathrm{mod}}) \approx \pi \Delta f \rho\, \delta(f_{\mathrm{mod}}) + \frac{\pi \rho}{4\Delta f}\,(\Delta f - f_{\mathrm{mod}}), \qquad \text{(A1)}$$

where Δf is the noise bandwidth, ρ is the power spectral density, and f_{mod} indicates modulation frequency. Besides the dc peak represented by the δ function, an approximately triangular continuous spectrum results. In the case of the squared envelope, the modulation spectrum has *exactly* a triangular shape besides the dc peak. This corresponds to the Wiener–Chintchin theorem which states that the Fourier transform of the squared signal equals the autocorrelation of the spectrum of the signal.

The following aspects are of particular relevance for modulation-detection experiments using a narrow-band noise as carrier: For a constant overall level of a noise band, the total power of intrinsic noise fluctuations, i.e., the total area under the triangle, remains constant. What changes is the spectral region over which the envelope spectrum stretches. Hence, with increasing noise bandwidth, the modulation spectrum becomes broader and flatter.

[1]This modification was motivated by physiological studies on adaptation in auditory nerve fibres where a comparable ratio of onset and steady state response was found (e.g., Smith and Zwislocki, 1975; Westerman and Smith, 1984). It was further assumed that the too strong overshoot at the output of the adaptation model in its original version (see Dau *et al.*, 1996a, b) would have a detrimental effect on psychoacoustical threshold predictions. However, the limitation of the onset response by Münkner (1993) was found to not have a significant influence on the results in the present study.

[2]The transfer functions of the resonance filters can be derived from the following recursive function: $y_i = e^{-\pi B \Delta} \cdot e^{-i2\pi f_0 \Delta} \cdot y_{i-1} + (1 - e^{-\pi B \Delta}) \cdot x_i$, where B is the filter bandwidth, f_0 is the center frequency of the resonance filter, and Δ is the inverse sample rate. The output y_i at time i depends on the input x_i at time i and on the last output value y_{i-1}.

[3]The adaptation loops transform fast envelope fluctuations nearly linearly. However, in the framework of the present model, without any further nonlinearity at a level where the signal envelope has already been extracted, it would not be possible to simulate a sufficient amount of *masking* in conditions with random modulation maskers. Particularly, in such masking experiments, the scaling of the modulation filters would not have an effect and masked thresholds would not depend on signal modulation frequency which is in contrast with experimental data (see also the accompanying paper by Dau *et al.*, 1997). A physiological motivation for the calculation of the envelope of the modulation filter output may be given by the finding of Langner and Schreiner (1988) that a much greater percentage of neurons in the central nucleus of the inferior colliculus of the cat show sensitivity for modulation *rate* than for modulation phase, indicating that at this stage of processing a modulation-rate place coding is performed and modulation phase information is reduced. Such a coding has already been incorporated by Hewitt and Meddis (1994) in a computer model of amplitude-modulation sensitivity of single units in the IC.

[4]Because of the relatively broad tuning of the modulation filters, some energy of a (stationary) signal also leaks into the transfer range of the overlapping modulation filters tuned to ''higher'' modulation frequencies. Thus the internal representation contains signal information in parallel at the output of several modulation filters, whereas in the original model version (Dau *et al.*, 1996a, b) only the lowest modulation channel (low-pass) contributed to the decision. Therefore, in the corresponding calibration experiment, a somewhat higher variance of the internal noise at the output of each modulation filter is required to satisfy the 1-dB criterion compared to the variance adjusted with the modulation low-pass approach described in Dau *et al.* (1996a, b).

[5]The optimal detector realized in the model clearly is an *application* of the original concept of the optimal detector developed in signal detection theory by Green and Swets (1966), in which—for the case of signal known exactly—the signal itself (and not the signal with noise) is used for the correlation with the received signal. It should be noted, as already mentioned in Dau *et al.* (1996a), that in actual masking experiments, the signal is typically not presented in isolation, and that, second, the presence of the ''masker'' influences the internal representation of the signal in a nonlinear way. It appears to be an appropriate strategy to extract the internal representation of the (normalized) template at a level well above threshold—comparable with the situation at the beginning of an actual experiment—containing just a small amount of internal and external noise (see, for example, the template from Fig. 6 in the present study). It is further noted that the current modeling approach realizes a decision device acting at the level of the internal representation of the stimuli. All information about the signal that is available at this stage of processing, is used in an ''optimal'' way. That is, information is combined optimally, although, for example, modulation phase is lost at a certain stage of preprocessing.

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking: Effects of modulation frequency, depth, and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Burns, E. M., and Viemeister, N. (**1981**). ''Played again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Creutzfeldt, O. D., Hellweg, F. C., and Schreiner, C. E. (**1980**). ''Thalamocortical transformation of responses to complex auditory stimuli,'' Exp. Brain Res. **39**, 87–104.

Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Doctoral thesis, University of Oldenburg.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.

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.

Eddins, D. (**1993**). ''Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region,'' J. Acoust. Soc. Am. **93**, 470–479.

Fassel, R. (**1994**). ''Experimente und Simulationsrechnungen zur Wahrnehmung von Amplitudenmodulationen im menschlichen Gehör,'' Doctoral thesis, University of Göttingen.

Fassel, R., and Kohlrausch, A. (**1995**). ''Modulation detection as a function of carrier frequency and level,'' IPO Annual Progress Report **30**, 21–29.

Fassel, R., and Kohlrausch, A. (**1996**). ''Sinusoidal amplitude modulation thresholds as a function of carrier frequency and level,'' J. Acoust. Soc. Am. **99**, 2566.

Fassel, R., and Püschel, D. (**1993**). ''Modulation detection and masking using deterministic and random maskers,'' in *Contributions to Psychological Acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), pp. 419–429.

Fassel, R., Kohlrausch, A., and Dau, T. (**1997**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' submitted to J. Acoust. Soc. Am.

Fleischer, H. (**1981**). ''Amplitudenmodulation von Terzrauschen: Experimente und theoretische Ergebnisse,'' Acustica **47**, 155–163.

Fleischer, H. (**1982a**). ''Modulationsschwellen von Schmalbandrauschen,'' Acustica **51**, 154–161.

Fleischer, H. (**1982b**). ''Calculating psychoacoustic parameters of amplitude modulated narrow noise bands,'' Biol. Cybern. **44**, 177–184.

Fleischer, H. (**1983**). ''Modulation thresholds of narrow noise bands,'' Proceedings of the 11th ICA, Paris 1983, pp. 99–102.

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.

Goldstein, J. L. (**1967**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.

Green, D. M. (**1973**). ''Temporal acuity as a function of frequency,'' J. Acoust. Soc. Am. **54**, 373–379.

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

Hartmann, W. M., and Pumplin, J. (**1988**). ''Noise power fluctuation and the masking of sine signals,'' J. Acoust. Soc. Am. **83**, 2277–2289.

Hewitt, M. J., and Meddis, R. (**1994**). ''A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus,'' J. Acoust. Soc. Am. **95**, 2145–2159.

Holube, I., and Kollmeier, B. (**1996**). ''Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model,'' J. Acoust. Soc. Am. **100**, 1703–1716.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

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,'' Acust. Acta Acust. **83**, 659–669.

Kollmeier, B., Dau, T., Hansen, M., and Holube, I. (**1996**). ''An Auditory-Model Framework for Psychoacoustics and Speech Perception and its Applications,'' Proceedings of the First Forum Acusticum, Antwerpen, published in Acust. Acta Acust. **82**, Suppl. 1, 89.

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

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).

Levitt, H. (**1971**). ''Transformed up–down procedures in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Maiwald, D. (**1967a**). ''Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen,'' Acustica **18**, 81–92.

Maiwald, D. (**1967b**). ''Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas,'' Acustica **18**, 193–207.

Martens, J. P. (**1982**). ''A new theory for multitone masking,'' J. Acoust. Soc. Am. **72**, 397–405.

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), pp. 251–308.

Münkner, S. (**1993**). ''Modellentwicklung und Messungen zur Wahrnehmung nichtstationärer Signale,'' Doctoral thesis, University of Göttingen.

Münkner, S., and Püschel, D. (**1993**). ''A psychoacoustical model for the perception of non-stationary sounds,'' in *Contributions to Psychological Acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), pp. 121–134.

Palmer, A. R. (**1982**). ''Encoding of rapid amplitude modulations by cochlear-nerve fibers in the guinea pig,'' Arch. Otorhinolaryngol. **236**, 197–202.

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

Püschel, D. (**1988**). ''Prinzipien der zeitlichen Analyse beim Hören,'' Doctoral thesis, University of Göttingen.

Ronken, D. A. (**1970**). ''Monaural detection of a phase difference in clicks,'' J. Acoust. Soc. Am. **47**, 1091–1099.

Schreiner, C., and Langner, G. (**1984**). ''Representation of periodicity information in the inferior colliculus of the cat,'' Soc. Neurosci. Abstr. **10**, 395.

Schreiner, C., and Urbas, J. V. (**1988**). ''Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields,'' Hearing Res. **32**, 49–65.

Sheft, S., and Yost, W. (**1990**). ''Temporal integration in amplitude modulation detection,'' J. Acoust. Soc. Am. **88**, 796–805.

Smith, R., and Zwislocki, J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.

Strickland, E. A., and Viemeister, N. F. (**1996**). ''Cues for discrimination of envelopes,'' J. Acoust. Soc. Am. **99**, 3638–3646.

Strickland, E. A., and Viemeister, N. F. (**1997**). ''The effects of frequency region and bandwidth on the temporal modulation transfer function'' J. Acoust. Soc. Am. **102**, 1799–1810.

Strube, H. W. (**1985**). ''A computationally efficient basilar-membrane model,'' Acustica **58**, 207–214.

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

Viemeister, N. F., and Plack, C. J. (**1993**). ''Time analysis,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 116–154.

Wier, C., Jesteadt, W., and Green, D. M. (**1977**). ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–184.

Westerman, L., and Smith, R. (**1984**). ''Rapid and short-term adaptation in auditory nerve responses,'' Hearing Res. **15**, 249–260.

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.

Zwicker, E. (**1952**). ''Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,'' Acustica **2**, 125–133.

Zwicker, E. (**1953**). ''Die Veränderung der Modulationsschwellen durch verdeckende Töne und Geräusche,'' Acustica **3**, 274–278.

Zwicker, E. (**1956**). ''Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs,'' Acustica **6**, 365–381.