Research Article

Masking Release for Sweeping Masker Components with Correlated Envelopes

JESKO L. VERHEY,¹ HENDRIKE KLEIN-HENNIG,² AND BASTIAN EPP^{1,3}

¹*Department of Experimental Audiology, Otto von Guericke University Magdeburg, Leipziger Str. 44, 31920 Magdeburg, Germany*

²*Institute of Physics, Carl von Ossietzky University Oldenburg, Carl von Ossietzky Str. 9-11, 26129 Oldenburg, Germany*

³*Department of Electrical Engineering, Centre for Applied Hearing Research, Technical University of Denmark, Ørstedes plads Building 352, 2800 Kongens Lyngby, Denmark*

Present address: Bastian Epp, Department of Electrical Engineering, Centre for Applied Hearing Research, Technical University of Denmark, Ørstedes plads Building 352, 2800 Kongens Lyngby, Denmark.

Received: 13 March 2012; Accepted: 11 September 2012; Online publication: 28 September 2012

ABSTRACT

To separate sounds from different sound sources, common properties of natural sounds are used by the auditory system, such as coherent temporal envelope fluctuations and correlated changes of frequency in different frequency regions. The present study investigates how the auditory system processes a combination of these cues using a generalized comodulation masking release (CMR) paradigm. CMR is the effect of a better signal detectability in the presence of comodulated maskers than in the presence of maskers with uncorrelated envelope fluctuations across frequencies. Using a flanking-band paradigm, the results of the first experiment of the present study show that CMR is still observed for the masker and the signal coherently sweeping up or down in frequency over time, up to a sweep rate of six octaves per second. Motivated by the successful modeling of CMR using filters sensitive to temporal modulations and recent physiological evidence of spectro-temporal modulation filters, the second experiment investigates whether CMR is also observed for spectro-temporal masker modulations generated using time-shifted versions of the masker envelope for each component. The thresholds increase as soon as the temporally coherent masker modulation is changed to a spectro-temporal masker modulation,

indicating that spectro-temporal modulation filters are presumably not required in CMR models.

Keywords: comodulation masking release, spectro-temporal modulations, auditory scene analysis

INTRODUCTION

A prominent property of many natural sounds, including speech, is the presence of correlated intensity fluctuations across frequencies (Nelken et al. 1999). Frequency components with coherent level fluctuations are likely to originate from one sound source. Likewise, a correlated change of the frequency in different frequency regions also indicates that these components belong to the sound of one sound source (Bregman 1994). The present study investigates whether the auditory system is able to use correlated intensity fluctuations as a cue to determine which stimulus components originate from one sound source when the stimulus components are frequency swept over time.

A psychoacoustical phenomenon indicating the sensitivity of the auditory system to a temporal coherence of intensity fluctuations across frequencies is comodulation masking release (CMR) (Hall et al. 1984; Verhey et al. 2003, for a review). Comodulation masking release is the effect of an enhanced ability to detect a sinusoid in the presence of a masker if the masker has the same

Correspondence to: Jesko L. Verhey · Department of Experimental Audiology · Otto von Guericke University Magdeburg · Leipziger Str. 44, 31920 Magdeburg, Germany. jesko.verhey@med.ovgu.de

temporal intensity fluctuations in different frequency regions, i.e., is comodulated. Comodulation masking release was initially shown in humans, but was also found in several other mammalian and non-mammalian species (Nelken et al. 1999; Pressnitzer et al. 2001; Nieder and Klump 2001; Klink et al. 2010; Branstetter and Finneran 2008).

The classical interpretation of CMR is that it reflects the ability of the auditory system to compare envelope fluctuations across frequencies (Verhey et al. 2003, for a review). A physiological correlate of CMR at the level of the cochlear nucleus indicates that wideband inhibition at this early level of the auditory pathway is a possible physiological mechanism underlying CMR (Pressnitzer et al. 2001; Neuert et al. 2004). At higher levels of the auditory system, envelope locking suppression was proposed as a physiological correlate of CMR (Nelken et al. 1999; Las et al. 2005). This envelope locking hypothesis is comparable to the model-based psychoacoustical hypothesis that a reduction of the modulation depth due to the presence of a target signal may be the cue used in the comodulated conditions (Verhey et al. 1999; Piechowiak et al. 2007). The psychoacoustical models of Verhey et al. (1999) and Piechowiak et al. (2007) are sensitive to a reduction of modulation depth by analyzing the temporal envelope fluctuations with a bank of overlapping bandpass filters, referred to as modulation filters. The assumption of a frequency tuning to temporal modulation was motivated by physiological findings in the inferior colliculus (IC) of the cat: Langner and Schreiner (1988) measured the response of units in the IC to sinusoidally amplitude-modulated tones with a frequency equal to the best frequency of the unit and various modulation frequencies and showed that some units are tuned in the temporal modulation frequency domain. Such a modulation tuning was found at several stages along the auditory pathway (Joris et al. 2004), and some studies suggest that, at the level of the cortex, there is a map encoding (envelope) periodicity orthogonal to the tonotopic representation (e.g., Langner et al. 2009; Baumann et al. 2011). Recent physiological studies indicate that the auditory system also shows a selectivity for spectro-temporal modulations, (e.g., deCharms et al. 1998; Depireux et al. 2001; Escabi and Schreiner 2002).

The first experiment of the present study is a flanking band experiment where the center frequency of each component is frequency swept over time. The aim of this experiment was to investigate whether a masking release can also be found with sweeping components.

The second experiment of the present study generalizes the paradigm of the first experiment from a temporally coherent masker modulation to a masker

modulation forming a complex spectro-temporal masker modulation pattern using time-shifted versions of the masker modulation for each band.

EXPERIMENT 1

Methods

Procedure. A three-alternative, forced-choice procedure with adaptive signal-level adjustment was used to determine the masked threshold of the sinusoidal signal. The intervals in a trial were separated by gaps of 500 ms. Subjects had to indicate which of the intervals contained the signal. The initial signal level was 80 dB SPL. Visual feedback was provided after each response. The signal level was adjusted according to a two-down, one-up rule to estimate the 70.7 % point of the psychometric function (Levitt 1971). The initial step size was 4 dB. After every second reversal, the step size was halved until a step size of 1 dB was reached. The run was then continued for another six reversals. The mean level at these last six reversals was used as an estimate of the threshold. The final individual threshold estimate was taken as the mean over four estimates.

Stimulus and apparatus. A common procedure for generating the narrow-band noise masker components $m_i(t)$ for a flanking-band experiment is to multiply a sinusoidal carrier with a low-pass-filtered noise.

$$m_i(t) = a_i \cdot \sin(2\pi \cdot f_i(t) \cdot t + \phi_i) \cdot N_i(t) \quad (1)$$

where $f_i(t)$ denotes the center frequency of the masker band, $N_i(t)$ denotes the low-pass-filtered noise without a DC component, and a_i and ϕ_i are the amplitude and the starting phase of the carrier, respectively. In the present study, a sweep was used as the carrier instead of a sinusoid, i.e.,

$$m_i(t) = a_i \cdot \sin\left(\int_0^t 2\pi \cdot f_i(t') dt' + \phi_i\right) \cdot N_i(t) \quad (2)$$

with the instantaneous frequency

$$f_i(t) = f_{i-\text{begin}} \cdot \left(\frac{f_{i-\text{end}}}{f_{i-\text{begin}}}\right)^{t/T} \quad (3)$$

where $f_{i-\text{begin}}$ denotes the initial frequency at stimulus onset, $f_{i-\text{end}}$ the frequency at stimulus offset, and T the stimulus duration in seconds. The sweeps were specified in units of octaves. Thus, a similar fraction of auditory filters are excited by each masker band. The starting phases, ϕ_i , were chosen randomly for each band. The amplitudes, a_i , were chosen such that each band had a level of 60 dB.

The relative frequency separation between the stimulus components was constant over time, i.e., the

same sweep rate was used for all stimulus components, since several previous studies on CMR showed that the frequency separation of the noise bands affects the magnitude of CMR (e.g., Schooneveldt and Moore 1987, 1989; Hall et al. 1990).

For the uncorrelated (UN) and comodulated (CM) conditions, the masker consisted of five of these noise bands. To generate a masker with uncorrelated masker bands (condition UN), different noise samples were used, i.e.,

$$M(t) = \sum_{i=1}^N a_i \cdot \sin\left(\int_0^t 2\pi \cdot f_i(t') dt' + \phi_i\right) \cdot N_i(t) \quad (4)$$

with

$$N_i \neq N_j \quad \text{for } i \neq j \quad (5)$$

Comodulated noise bands (condition CM) were generated using the same low-pass-filtered noise, $N(t)$, for all masker components:

$$M(t) = \sum_{i=1}^N a_i \cdot \sin\left(\int_0^t 2\pi \cdot f_i(t') dt' + \phi_i\right) \cdot N(t) \quad (6)$$

Each noise band had a bandwidth of 24 Hz and a duration of 500 ms, including 50-ms raised-cosine ramps at masker onset and offset.

In addition to the conditions UN and CM, thresholds were also measured for the masker that only consisted of the signal-centered band (SCB). This condition was included to measure the ability of the auditory system to detect the signal when masker and signal had a similar frequency content. In the following, it will be referred to as the SCBonly condition.

Thirteen sweep rates, r , were used: 0 (noSweep condition), $\pm 2/3$, ± 2 , ± 3 , ± 4 , ± 5 , and ± 6 octaves per second (sweep conditions). The sweep rate is defined as the temporal rate of change in center frequency of one masker band. Positive sweep rates correspond to upward sweeps in frequency and negative sweep rates to downward sweeps in frequency. For the noSweep condition, the center frequency was 1,000 Hz for the SCB and 400, 600, 1,400, and 1,600 Hz for the flanking bands. For the other (sweep) conditions, the frequency $f_{i-\text{begin}}$ was chosen such that (1) the sweep started below and ended above the center frequencies of the corresponding masker bands in the noSweep condition and (2) the instantaneous frequencies of the corresponding bands were equal to the center frequencies for the noSweep condition at $t = T/2$.

$$f_{i-\text{begin}} = f_{i-\text{noSweep}} \cdot 2^{-r \cdot T/2} \quad (7)$$

Figure 1 shows the spectrograms of different masker realizations in the masker conditions UN

(left) and CM (right) for sweep rates of -4 , 0 , and $+4$ octaves per second. The signal can be described by the following expression:

$$S(t) = a_s \cdot \sin\left(\int_0^t 2\pi \cdot f_s(t') dt'\right) \quad (8)$$

where a_s is the amplitude and $f_s(t')$ is the instantaneous frequency of the signal. For the noSweep condition, the frequency was 1,000 Hz. For the sweep conditions, the sweep rate was the same for signal and the masker bands, i.e., for each instant in time, the SCB was centered at the signal frequency. The signal duration was equal to the masker duration, i.e., 500 ms. The signal was gated on and off with 50-ms raised-cosine ramps. All stimuli were generated digitally with a sampling frequency of 44,100 Hz using MATLAB. Signals were converted to an analog signal (RME ADI-8 DS), amplified (Tucker Davis Technologies TDT HB7), and presented diotically to the listeners in a double-walled sound-attenuating booth via headphones (Sennheiser HD 650).

Listeners. Eight listeners participated. None of the listeners had any history of hearing difficulties, and their thresholds in quiet were within 15 dB HL for all standard audiometric frequencies between 250 Hz and 4,000 Hz. The listeners had at least 1 h experience in experiments on CMR before collecting the data. The study was approved by the ethical committee of the University of Oldenburg.

Statistical analysis. To examine the statistical significance of the effects of stimulus variables and interindividual differences, a three-way analysis of variance (ANOVA) was performed (Mathworks MATLAB, version R2010b, and MATLAB statistics toolbox). To explore the sources of significant effects, a multiple comparison test with Bonferroni correction was performed (MATLAB statistics toolbox). A significance level of 0.05 was used as a criterion.

Results

The results are shown in Figure 2. In the upper panel, the mean thresholds of the signal for the UN, CM, and SCBonly conditions are plotted in decibels SPL as a function of the sweep rate. The CMR, i.e., the difference between the corresponding thresholds for conditions UN and CM, are shown in the lower panel. For a sweep rate of 0 octaves per second, the thresholds are 56 dB SPL for condition UN and about 45 dB SPL for condition CM. This results in a CMR of 11 dB. With increasing magnitude of the negative sweep rate, the thresholds increase for both conditions. For a sweep rate of -6 octaves per second, the thresholds are 64 dB SPL for condition UN and about 58 dB SPL

FIG. 1. Spectrograms of one masker realization for each condition as used in the first experiment. The *upper*, *middle*, and *lower panels* show the masker with sweep rates of -4 octaves per second, 0 octaves per second (noSweep condition), and the $+4$ octaves per second. The *left* and *right columns* show different realizations of the maskers for the uncorrelated (UN) and the comodulated (CM) conditions, respectively. The spectrograms were computed using Hamming windows with a length of 1,024 samples (about 23 ms) and 50 % overlap.

for condition CM, i.e., the CMR decreases to about 6 dB. The data for positive sweep rates are similar to those for negative sweep rates. The thresholds increase to 65 dB SPL for a sweep rate of 6 octaves per second for conditions UN and to about 59 dB SPL for condition CM. CMR decreases to about 5 dB. Thus, the CMR for large absolute sweep rates is smaller than the CMR for the noSweep condition. However, it is still clearly present, indicating that, even at high rates, comodulation can be used as a cue. For the condition SCBonly, the effect of sweep rate is less

pronounced than for the conditions CM and UN. For the sweep rate of 0 octaves per second, the thresholds are the same as for the UN condition, indicating no energetic masking due to the addition of the flanking bands. The thresholds increase slightly toward negative sweep rates from 55 dB SPL for a sweep rate of 0 octaves per second to 59 dB SPL for a sweep rate of -6 octaves per second.

The statistical significance of the effects was performed with the factors listener(A) $\times$ sweep rate(B) $\times$ condition(C). The stimulus variables sweep rate (13 levels: ± 6 , ± 5 , ± 4 , ± 3 , ± 2 , $\pm 2/3$, 0) and condition (three levels: SCBonly, UN, CM) were fixed factors. The listener (eight levels: L1–L8) was a random factor. The ANOVA revealed significant effects of all factors (A: $F_{(7,816)}=7.16$, B: $F_{(12,995)}=45.92$, C: $F_{(2,7744)}=78.44$) and interactions (A $\times$ B: $F_{(84,22)}=3.31$, A $\times$ C: $F_{(14,99)}=15.09$, B $\times$ C: $F_{(24,159)}=24.25$). The post hoc multiple comparison test revealed that the thresholds in condition SCBonly are significantly different for sweep rates of -4 , -5 , and -6 octaves per second compared to the noSweep condition. The thresholds in condition UN are significantly different for sweep rates of ± 3 , ± 4 , ± 5 , and ± 6 octaves per second compared to the noSweep condition. The thresholds for condition CM are significantly different for sweep rates of ± 2 , ± 3 , ± 4 , ± 5 , and ± 6 octaves per second compared to the noSweep condition. At each level of the sweep rate, the thresholds for conditions UN and CM are significantly different from each other. The thresholds for conditions SCBonly and CM only differ significantly for sweep rates smaller than ± 5 octaves per second. Thresholds of the same condition with the same absolute sweep rate do not differ significantly between the upward- and downward-sweeping components.

FIG. 2. Mean thresholds over all listeners for the sweep rates used in the first experiment. The *upper panel* shows the detection thresholds for the conditions UN (squares) and CM (triangles). In addition, the thresholds for the SCBonly are shown with crosses connected with a gray line. The *lower panel* shows the resulting CMR. The error bars indicate inter-individual standard errors.

These data indicate the ability of the auditory system to make use of comodulation even if the stimulus components are swept in frequency.

EXPERIMENT 2

Methods

The experimental setup, procedure, listeners, and statistical analysis were the same as for the previous experiment. Four spectro-temporal masker modulation conditions were used with a fixed carrier sweep rate of +2/3 octaves per second. A spectral component in the modulator of the masker was introduced by temporally shifting the masker bands relative to each other in such a way that the ratio of the spectral distance between the bands and their relative time shift was the same for all combinations of masker bands. A similar approach to generate spectro-temporal modulations was used in Chi et al. (1999) to measure detection thresholds for these modulations in humans. Chi et al. (1999) used starting phases for the sinusoidal modulation that increased (or decreased) linearly with the carrier component which were equally spaced on a logarithmic frequency axis. These time phase shifts introduced a spectral modulation component in the modulation. For their modulator, these modulations were also sinusoidal on a log-frequency axis (see their Fig. 3). A phase shift for a sinusoid is equivalent to a time delay which is used in the present study. In contrast to Chi et al. (1999), the present study used narrowband noise and the carriers are frequency sweeps instead of sinusoids, but otherwise, the approach is the same as in their study. The chosen time shifts generate a certain slope, μ , in the envelope pattern, as visualized in Figure 3 with the dotted line. The absolute value of this slope is equal to the ratio of

the frequency distance and relative time shift minus the carrier sweep rate. For the chosen time shifts, the slope μ was approximately equal to -3, -7, -37, and -75 octaves per second. A negative slope implies that the time shifts decrease as a function of the initial frequency of the components. In addition, two conditions from the first experiment are included for comparison: condition CM with a slope μ of ∞ octaves per second, i.e., for temporally aligned bands, and condition UN with a slope of 0 octaves per second, i.e., uncorrelated bands.

The procedure for generating the masker was essentially the same as for the first experiment. To allow for time shifts, the noise bands were generated with a duration of 2,000 ms. A spectral modulation component was introduced into the masker modulation by choosing different portions of the masker with temporal delays, τ_i , for the flanking band.

$$M(t) = \sum_{i=1}^N a_i \cdot \sin\left(\int_0^t 2\pi \cdot f_i(t') dt' + \phi_i\right) \cdot N_i(t + \tau_i). \quad (9)$$

The corresponding time shifts are given in Table 1.

The masker was gated with a window of 500-ms duration including 50-ms raised-cosine ramps at masker onset and offset. The signal was the same as in the first experiment, with the sweep rate of +2/3 octaves per second.

Results

The results are shown in Figure 4. The mean thresholds are plotted in decibels SPL for the different masker conditions. The threshold for the comodu-

FIG. 3. Spectrograms of one masker realization for each slope of the spectro-temporal masker modulation used in the second experiment. **A** Masker with a temporal alignment of the masker bands with a slope of 1 octaves per second (condition CM of first experiment). **B–E** Masker with temporally shifted masker bands with slopes, μ , of -75, -37, -7, and -3 octaves per second. **F** Masker with uncorrelated masker bands (condition UN of the first experiment). The dashed lines indicate the slopes.

TABLE 1

Time shifts, τ_i , used to introduce a spectral component in the masker modulation

Slope	Time shift, τ_i (ms)				
	"1"	" $\mu=-75$ "	" $\mu=-37$ "	" $\mu=-7$ "	" $\mu=-3$ "
$f_{i-noSweep}$ (Hz)					
400	0	-17.4	-34.9	-174.4	-349
600	0	-9.7	-19.5	-97.3	-195
1,000	0	0	0	0	0
1,400	0	6.4	12.8	64.1	128.1
1,600	0	9	17.9	89.5	179

The time shifts are given for each masker band identified by the center frequency in the noSweep condition

lated masker with a slope of ∞ octaves per second (CM) is shown as a triangle and the threshold for the uncorrelated masker (UN) shown as a square. The thresholds for the masker conditions with slopes of -75 , -37 , -7 , and -3 octaves per second are shown as diamonds. The threshold for condition ∞ (CM) from the first experiment is about 45 dB SPL and the threshold for condition UN about 55 dB SPL. The thresholds for the other conditions increase with decreasing magnitude of the slope from about 50 dB SPL to about 57 dB SPL. Thus, as soon as a spectral modulation component is introduced, the CMR decreases.

Statistical analysis was carried out with the factors listener(A) $\times$ slope(B). The stimulus variable slope (six levels: ∞ (CM), -75 , -37 , -7 , -3 , UN) was a fixed factor. The listener (eight levels: L1–L8) was a random factor. The ANOVA revealed a significant effect of the two factors (A: $F_{(7,61)}=10.35$, B: $F_{(5,591)}=101$), but not of the interaction (A $\times$ B: $F_{(35,6)}=1.42$). The post hoc multiple comparison test revealed that the threshold in condition UN is significantly different compared to the conditions with slopes of -3 , -37 , and -75 and ∞ (CM).

FIG. 4. Mean thresholds over all listeners for the slopes used in the second experiment and for uncorrelated maskers. The threshold for a slope of 1 octaves per second (CM) is shown by a triangle. The thresholds for the temporally shifted maskers with different slopes are shown by diamonds. The threshold for the uncorrelated masker is shown as a square. Error bars indicate inter-individual standard errors.

DISCUSSION

A few previous studies investigated how a small sinusoidal frequency modulation (FM) of the masker bands affects CMR (Grose and Hall 1990). The maximum peak deviation of these sinusoidal FMs was only 6 % of the center frequency. They measured no effect of FM modulation coherence on the magnitude of CMR for a sinusoidal signal embedded in FM-modulated noise bands. The CMR was about the same for 3 and 6 % FM modulation.

In contrast to those studies, the center frequencies of the masker bands and the signal of the present study changed monotonically over time, and sweep rates of up to several octaves per second were used such that the masker excited different auditory filters during the stimulus period. The data of the first experiment indicate that coherence of the masker envelopes is beneficial for signal detection of the signal up to high sweep rates. For all conditions, masked thresholds increase as the magnitude of the sweep rate increases.

The second experiment shows that the CMR decreases as soon as a temporal masker modulation is changed to a spectro-temporal masker modulation.

Role of energy within an auditory filter and across-frequency integration of information

An increase in threshold as the magnitude of the sweep rate is increased is common to all conditions of the first experiment. To test to what extent this increase can already be accounted for by the energy in one auditory filter, a power spectrum model (PSM) was used (Fletcher 1940) to predict the data for conditions UN and SCBonly. Condition CM was not simulated since the PSM is not sensitive to comodulation, i.e., cannot distinguish between the conditions UN and CM.

The signal was filtered by a fourth-order gamma-tone filter (Hohmann 2002) centered at 1,000 Hz, i.e., the signal frequency for the noSweep condition. The stimuli were presented in the same way as in psychoacoustical measurements, i.e., thresholds were determined with a three-interval forced-choice procedure with adaptive level tracking (two-down, one-up).

The root mean square (RMS) of the filter output excluding onset and offset was used to estimate the level at the output of the filter. The resulting level in the interval where both signal and noise were present was compared to the maximum levels of the intervals where the noise only was present. If the difference exceeded 1 dB, the signal was assumed to be detected.

Figure 5 shows the mean simulation results over 100 runs with the PSM as an artificial observer. The symbols are the same as in Figure 2. Simulated

FIG. 5. Predicted thresholds for the conditions UN and SCB only of the first experiment. For the simulation, the power spectrum model was used. Error bars indicate standard deviations of the simulated runs.

thresholds are lowest for stationary maskers, i.e., a sweep rate of 0 octaves per second. At this sweep rate, the simulated thresholds for conditions UN and SCB only are about the same, i.e., the addition of flanking bands did not increase the masker energy in the filter centered at the signal frequency. With increasing sweep rates, the thresholds for condition SCB only increase from about 55 dB SPL for a sweep rate of 0 octaves per second to about 63 dB SPL for the highest sweep rates. This increase is due to the reduced duration of the stimuli in the passband of the filter and the increased influence of masker fluctuations on the estimate of the energy in each interval (see deBoer 1966 for a theoretical discussion on the influence of these stimulus parameters on signal detection). The predicted increase is considerably higher than in the data. This indicates that signal detection is based on the analysis of more than the output of one auditory filter. This is reasonable since the sweeping signals with high sweep rates excite more than one filter. Increasing the sweep rate increases the number of auditory channels containing the signal. If it is assumed, for simplicity, that the auditory filters are third-octave filters, then the stimulus excites nine filters during its duration. Each of these filters contains a ninth of the total stimulus duration. Assuming further that the information is integrated across the output of auditory filters containing the stimulus (Green 1958), the thresholds would be predicted to be about $5 \text{ dB} (10 \cdot \log_{10}(\sqrt{9}))$ lower for a sweep rate of 6 octaves per second than those predicted with the PSM using only the output of the auditory filter centered at 1,000 Hz. The resulting predicted increase of 3 dB is in good agreement with the measured effect (2 dB for positive sweep rates and 4 dB for negative sweep rates).

Using only the output of the auditory filter centered at 1,000 Hz, the PSM-predicted thresholds for condition UN increased from about 55 dB SPL to about 66 dB SPL for the highest sweep rates (squares in Fig. 5). The higher predicted thresholds for condition UN than for the SCB only masker are due to the additional masker

energy: As the sweep rate increases, the masker energy within the filter increases since more and more masker energy of the flanking bands falls into the passband of the auditory filter, i.e., the signal-to-noise ratio at the output of the filter decreases. The predicted increase of 11 dB is slightly larger than the 8–9 dB observed in the data. The difference is presumably due to the restriction of the model to analyze the auditory filter centered at 1,000 Hz. However, the difference is considerably smaller than the 5 dB, which would be expected on the basis of an integration of information across the output of auditory filters. This indicates that this integration of information is less effective when flanking bands are present. The higher similarity in the slope of the data for the conditions UN and CM than for the conditions CM and SCB only supports this hypothesis.

Role of temporal modulation processing in an auditory filter

Recently, Hsieh and Saberi (2010) showed that detection of amplitude modulation is impaired when imposed on a sweeping carrier instead of a stationary tonal or noise carrier. Thus, if the masking release for a comodulated masker is due to the analysis of the modulation spectrum, as suggested by Verhey et al. (1999) and Piechowiak et al. (2007), the decrease in CMR with increasing sweep rate in the first experiment may be partly due to the reduced ability to use modulation cues for sweeping components. In addition, it may be difficult to detect common modulation for the different masker bands. The classical view of across-channel processes in CMR is that the auditory system gains information about the masker envelope in the auditory filter from the off-frequency auditory filters centered at the flanking bands (see Verhey et al. 2003 for a review). Without the signal, the envelopes in the filters are highly correlated. This correlation is reduced when the signal is added. The present data challenge this view, as visualized in Figure 1. Thus, a process based on the correlation of the envelopes at the output of each auditory filter has difficulties in detecting a reduction in comodulation due to the addition of the signal for large sweep rates since envelope correlation is already considerably reduced for the masker alone. Figure 6 shows the time signals (B,D) and the envelopes (A,C) of the output of auditory filters modeled by fourth-order gammatone filters (Hohmann 2002) for two comodulated maskers. The envelopes were calculated by a Hilbert transform of the signals at the output of the filters. Panel A shows the envelopes of the filter outputs for a comodulated masker with a sweep rate of 0 octaves per second. The time signal and the spectrogram of the masker are shown in panel B and in the middle-right panel of Figure 1, respectively.

FIG. 6. Envelopes in different auditory channels modeled by a fourth-order gammatone filter bank for the comodulated maskers with a sweep rate of 0 octaves per second (A) and a sweep rate of 4 octaves per second (C). The noise samples (B, D) were the same as those used to generate the spectrograms in the *middle* and *bottom right* panels of Figure 1.

The envelope is highly correlated across filters, as expected for a comodulated masker. Panel C shows the corresponding envelopes for the masker with a sweep rate of 4 octaves per second (bottom right panel of Fig. 1). Thus, a process analyzing the masker modulation at the output of each auditory filter has difficulties in detecting a reduction in comodulation due to the addition of the signal for large sweep rates. The correlation is considerably reduced, i.e., it is unlikely that a process as described above is used to detect the signal. In contrast to this prediction, the data only show a gradual decrease in CMR, i.e., the auditory system is only slightly impaired when the components are swept in frequency over time. The gradual decrease of CMR with increasing sweep rates supports the hypothesis that the auditory system uses a stage with a broader frequency selectivity than the peripheral auditory filter to detect comodulation. An example for such a broadly tuned stage which has been investigated using electrophysiology is the wide-band inhibitor (Pressnitzer et al. 2001; Neuert et al. 2004). A slightly different broadly tuned stage was proposed in the model by Piechowiak et al. (2007) where the output of modulation filters are combined across auditory filters. Such mechanism may also detect the comodulation in conditions with sweeping components, at least for slow-amplitude modulations where the modulation filters preserve the temporal waveform of the envelope.

Role of spectro-temporal modulation processing

An analysis of temporal modulation with a modulation filterbank has been shown to be a sufficient mechanism to predict CMR on the basis of modulation cues (Verhey et al. 1999; Verhey and Ernst 2009; Piechowiak et al. 2007). The present study investigated whether frequency selectivity for spectro-temporal modulations, as observed in physiological studies (e.g., deCharms et al. 1998; Depireux et al. 2001; Escabi and Schreiner 2002), may also be used by the auditory system to detect signals in complex spectro-temporal masker modulations. The results of the second experiment of the present study indicate a dominance of temporal modulation cues in CMR since the masking release decreased as soon as the purely temporal masker modulation was changed to a spectro-temporal masker modulation. The finding is in qualitative agreement with the psychoacoustical results of previous CMR studies if the effect of the spectral component in the spectro-temporal masker modulation is interpreted as a time delay of the bands. For a single flanking band, McFadden (1986) and Moore and Schooneveldt (1990) found that the CMR started to diminish when a small time delay was introduced, in agreement with the data of the present study. This indicates that spectral masker modulation components combined with temporal masker modulations are not beneficial for signal detection in noise or, at least, that processing of temporally coherent amplitude fluctuations is much more important in this task.

SUMMARY AND CONCLUSIONS

Comodulation masking release (CMR) has so far only been investigated using stationary spectra or using spectra with frequency modulations within a small frequency range. The present study shows that CMR also occurred for sweeping masker and signal components, up to a sweep rate of 6 octaves per second, if the CMR was quantified relative to the uncorrelated condition, i.e., when the two conditions with the same number of masker bands were compared.

The results for the masker consisting of a signal-centered band (SCB) only show a trend different from the other maskers. Simulations with a power spectrum model showed that the thresholds for the condition SCBonly can be predicted by assuming an energy-based detection within an auditory channel in combination of an integration of information across frequencies. Predictions of the thresholds for the uncorrelated (UN) condition indicate that this integration is less effective when flanking bands are present.

The data support the hypothesis that the auditory system uses a broadly tuned stage to detect coherent envelope fluctuations in different frequency regions rather than a separate analysis of the envelope of the output of each auditory filter.

While CMR is found for non-stationary spectra, it seems crucial that the CMR is only obtained for purely temporal masker modulations. The CMR is reduced as soon as a purely temporal masker modulation is changed to a spectro-temporal masker modulation. This result suggests that comodulation and spectro-temporal patterns are not processed by the same mechanism and are treated as different perceptual processing dimensions, contributing independently to the decision which frequency components are likely to arise from the same sound source.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB/TRR 31).

REFERENCES

- BAUMANN S, GRIFFITHS TD, SUN L, PETKOV C, THIELE A, REES A (2011) Orthogonal representation of sound dimensions in the primate midbrain. *Nat Neurosci* 14:423–425
- BRANSTETTER BK, FINNERAN JJ (2008) Comodulation masking release in bottlenose dolphins (*Tursiops truncatus*). *J Acoust Soc Am* 124:625–633
- BREGMAN, A. S. (1994) Auditory Scene Analysis. The Perceptual Organization of Sound. MIT Press, Cambridge
- CHI TS, GAO YJ, GUYTON MC, RU PW, SHAMMA S (1999) Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106:2719–2732
- DEBOER E (1966) Intensity discrimination of fluctuating signals. *J Acoust Soc Am* 40:552–560
- DECHARMS RC, BLAKE DT, MERZENICH MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439–1443
- DEPIREUX DA, SIMON JZ, KLEIN DJ, SHAMMA SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophys* 85:1220–1234
- ESCARI MA, SCHREINER CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci* 22:4114–4131
- FLETCHER H (1940) Auditory patterns. *Rev Mod Phys* 12:47–61
- GREEN DM (1958) Detection of multiple component signals in noise. *J Acoust Soc Am* 30:904–911
- GROSE JH, HALL JW (1990) The effect of modulation coherence on signal threshold in frequency-modulated noise bands. *J Acoust Soc Am* 88:703–710
- HALL JW, HAGGARD MP, FERNANDES MA (1984) Detection in noise by spectro-temporal pattern analysis. *J Acoust Soc Am* 76:50–56
- HALL JW, GROSE JH, HAGGARD MP (1990) Effects of flanking band proximity, number, and modulation pattern on comodulation masking release. *J Acoust Soc Am* 87:269–283
- HOHMANN V (2002) Frequency analysis and synthesis using a gammatone filterbank. *Acta Acoustica/Acustica* 88:433–442
- HSIEH IH, SABERI K (2010) Detection of sinusoidal amplitude modulation in logarithmic frequency sweeps across wide regions of the spectrum. *Hear Res* 262:9–18
- JORIS P, SCHREINER CE, REES A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541–577
- KLINK KB, DIERKER H, BEUTELMANN R, KLUMP GM (2010) Comodulation masking release determined in the mouse (*Mus musculus*) using a flanking-band paradigm. *J Assoc Res Otolaryngol* 11:79–88
- LANGNER G, SCHREINER CE (1988) Periodicity coding in the inferior colliculus of the cat: I. Neuronal mechanisms. *J Neurophys* 60:1799–1822
- LANGNER G, DINSE HR, GODDE B (2009) A map of periodicity orthogonal to frequency representation in the cat auditory cortex. *Front Integr Neurosci* 3:1–11
- LAS L, STERN E, NELKEN I (2005) Representation of tone in fluctuating maskers in the ascending auditory system. *J Neurosci* 25:1503–1513
- LEVITT H (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477
- McFADDEN D (1986) Comodulation masking release: effects of varying the level, duration, and time delay of the cue band. *J Acoust Soc Am* 80:1658–1667
- MOORE BCJ, SCHOONEVELDT GP (1990) Comodulation masking release as a function of bandwidth and time-delay between on-frequency and flanking-band maskers. *J Acoust Soc Am* 88:725–731
- NELKEN I, ROTMAN Y, YOSEF O (1999) Responses of auditory cortex neurons to structural features of natural sounds. *Nature* 397:154–157
- NEUERT V, VERHEY J, WINTER I (2004) Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers. *J Neurosci* 24:5789–5797
- NIEDER A, KLUMP G (2001) Signal detection in amplitude-modulated maskers. II. Processing in the songbird's auditory forebrain. *Eur J Neurosci* 13:1033–1044
- PIECHOWIAK T, EWERT SD, DAU T (2007) Modeling comodulation masking release using an equalization–cancellation mechanism. *J Acoust Soc Am* 121:2111–2126
- PRESSNITZER D, MEDDIS R, DELAHAYE R, WINTER I (2001) Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. *J Neurosci* 21:6377–6386
- SCHOONEVELDT GP, MOORE BCJ (1987) Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band. *J Acoust Soc Am* 82:1944–1956
- SCHOONEVELDT GP, MOORE BCJ (1989) Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking-bands. *J Acoust Soc Am* 85:262–272
- VERHEY JL, ERNST SMA (2009) Comodulation masking release for regular and irregular modulators. *Hear Res* 253:97–106
- VERHEY JL, DAU T, KOLLMEIER B (1999) Within-channel cues in comodulation masking release (CMR): experiments and model predictions using a modulation filter-bank model. *J Acoust Soc Am* 106:2733–2745
- VERHEY J, PRESSNITZER D, WINTER I (2003) The psychophysics and physiology of comodulation masking release. *Exp Brain Res* 153:405–417