CARL VON OSSIETZKY universität OLDENBURG

Bachelor thesis

Across frequency processing with time varying spectra

Handed in by Hendrike Heidemann

Study course: Engineering Physics

First supervisor: Prof. Dr. Jesko Verhey

Second supervisor: Prof. Dr. Volker Mellert

Oldenburg, 18.07.2008

Contents

1. Introduction

Humans have the ability to orient themselves in acoustically complex situations. One important prerequisite for this is the ability to distuingish between sounds that contain important information and sounds that are useless and can be considered background noise. A possibility to investigate this ability is to measure the detectibility of a tone within noise. The lowest level for which the tone is heard is called the masked threshold. The detectability of a tone masked by narrow band noise can be increased by adding noise bands with a time synchronous amplitude modulation outside the auditory filter of the signal. The reduction in threshold is called comodulation masking release (CMR). Sections 2.2 and 2.3 will give a short overview about this effect and the possible underlying physiological principles.

CMR experiments are well researched for stimuli with temporally constant spectra. This work investigates CMR effects for stimuli with time varying spectra. The first question, which is discussed in chapter 3 is whether CMR still occurs if signal and masking bands are sweeped. Sweeped stimuli can cover a frequency range of several auditory filters. An interesting question is whether humans are able to combine information about the signal out of several auditory filters or whether the output of only one filter is evaluated. Wide band inhibition (see chapter 2.3) serves as a possible physiological explanation for CMR. It has to be discussed whether the thresholds for sweeped stimuli can be explained in the same way.

The second experiment deals with the question, if it is necessary to have a time coherent comodulation as used in previous studies and in the first experiment of the present study. As will be described in section 2.4 single neurons correspond best to spectro-temporal modulations created by ripple stimuli. Ripple stimuli do not have time coherent modulations, but spectro-temporal modulations. Chapter 4 deals with the question whether spectro-temporal modulations also cause a release in threshold similar to CMR.

2. Basic principles

2.1. Auditory filters

Soundwaves reaching the ear have to be transformed to neural impulses in order to be interpreted by the brain. This transformation takes place on the basilar membrane in the inner ear. Each frequncy component of a soundwave excites vibrations of the basilar membrane on a certain position. Low frequencies are mapped onto the end of the basilar membrane. High frequencies are mapped onto the beginning of the basilar membrane close to the oval window [5]. The basilar membrane can be modelled by a filterbank of overlapping bandpass filters. Filters become wider for higher frequencies. The shape of each filter is asymmetric with a steeper slope towards higher frequencies as shown in figure 1.

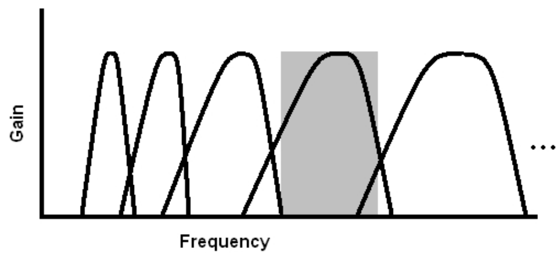


Figure 1: Bank of auditory filters. The grey rectangle indicates the equivalent rectangular filter belonging to the fourth auditory filter.

A simple way to describe the bandwidth of a filter is to use its equivalent rectangular bandwidth. The equivalent rectangular bandwidth is the bandwidth which is needed, such that the same amount of power is collected by a rectangular filter as by the real filter. A good approximation is an equivalent rectangular bandwidth of 1/3 octave [8].

2.2. CMR

Many psychoacoustic experiments deal with the question under which circumstances humans are able to detect a signal masked by noise. The detectability of a tone masked by noise depends, in addition to the signal to noise ratio, on various parameters of the noise. One important parameter is the noise statistics. The threshold for the signal masked by a noise band centered at the signal frequency can be lowered by adding noise bands outside the auditory filter. A reduction of the threshold is achieved, if the additional noise bands have the same temporal envelope as the masking band. Noise bands fulfilling these requirements are called comodulated. This effect of threshold reduction is called comodulation masking release (CMR). In experiements dealing with CMR three masker conditions are compared to each other. In the reference condition (RF) only one narrow noise band centered at the signal frequency is present. In the comodulated condition (CM) one or more comodulated noise bands at different frequencies are added. The third condition is the uncorrelated condition (UN). In this condition the same number of noise bands are present as in the comodulated condition, with the difference that the bands are uncorrelated. Uncorrelated noise bands have independent envelopes. The threshold for the comodulated condition is usually lower than for the reference condition and the uncorrelated condition. There are two possiblities to define CMR. One compares the single masking band, called on frequency masker (OFM), with the situation in which additional comodulated flanking bands (FB) are present. The resulting threshold reduction is indicated as CMR_{RF-CM}. Another way to define CMR is to relate the threshold obtained in a comodulated situation to the threshold of the uncorrelated situation. The resulting threshold reduction is indicated as CMR_{UN-CM}. The three conditions are shown in figure 2.

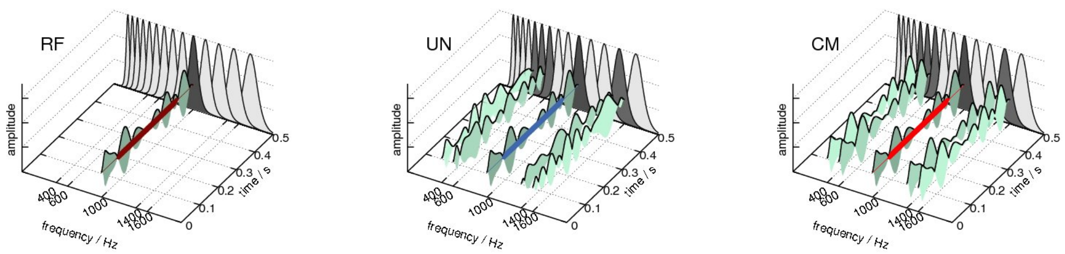


Figure 2: Three conditions of a comodulation masking release (CMR) flanking band (FB) experiment with 4 FBs and the on-frquancy masker (OFM)including the signal. Reference condition (RF): only the narrow band OFM, uncorrelated condition (UN): OFM and 4 FBs containing different envelopes, comodulated condition (CM): OFM and FBs have the same envelope

The threshold reduction in comodulated conditions versus reference and uncorrelated conditions depends on different stimuli parameters. Parameters enhancing the amount of CMR are: a large masker bandwidth, low modulation frequency, high modulation depth, a regular envelope and a high masker spectrum level [14]. The reduction in threshold for the comodulated situation is interpreted as an across-frequency process in which the auditory system is able to compare the envelope from different auditory filters. In the comodulated situation the envelopes of the noise bands are the same. The signal slightly changes the envelope of the on-frequency masker, which now differs from the envelopes of the other noise bands. By comparing the evelopes from different auditory filters, this difference can be detected even at low signal to noise ratios. There are several models trying to explain CMR. The dip listening model suggests that the decrease in threshold can be explained by comparing the envelopes at the times for which noise energy is low. For these times the signal alters the envelope significantly. This difference can be detected. Another important model which tries to explain CMR is the equalization cancellation model. It suggests that in a first stage, the equalization stage, the envelopes of different flanking bands are shifted in time. The amplitude is adjusted to match the envelope of the on-frequency masker (OFM). In a second stage the resulting envelopes are subtracted from the on-frequency masker and only the signal remains [12]. There are models which explain CMR without using any across-channel information. Instead they assume that in the comodulated condition a reduction of modulation depth caused by the signal can be detected [10]. However little is known about the underlying neural principles. One hypothesis to explain CMR on a physiological basis is wide band inhibition.

2.3. Wide band inhibition

One hypothesis states that not only the frequency information is encoded in the neurons of the auditory path, but also information about the envelope can be transmitted [14]. The acoustic signal is split up and processed according to its frequency distribution in one or several auditory filters. In each filter separate neurons encode the signal information. Two types of neurons, playing a large role trying to explain CMR in the cochlear nucleus, can be distinguished. The first type is the so-called chopper cell. Chopper cells are tuned to one narrow frequency range and their spike rate is the image of the structure of the stimulus. During times with a high filter input their firing rate is higher than in times with a low filter input. The second class of cells are called onset neurons. They cover a wide frequency range and fire at the beginning of a stimulus. After that the firing rate of onset neurons represents the structure of wide band stimulus. Narrow band signals such as tones have little effect on their response. The responses of the two neurons are presented in figure 3.

If one assumes an inhibitory effect of the onset neurons on the chopper cells, the spike rate of the chopper cells is reduced during times of high masker energy. The spike rate during times with a good signal to noise ratio is not influenced by the inhibitory effect of the onset cells. The interaction between the two cell types is called wide band inhibition. The model shown in figure 4 clarifies the process. In this model the chopper neurons are abbreviated by NB for narrow band

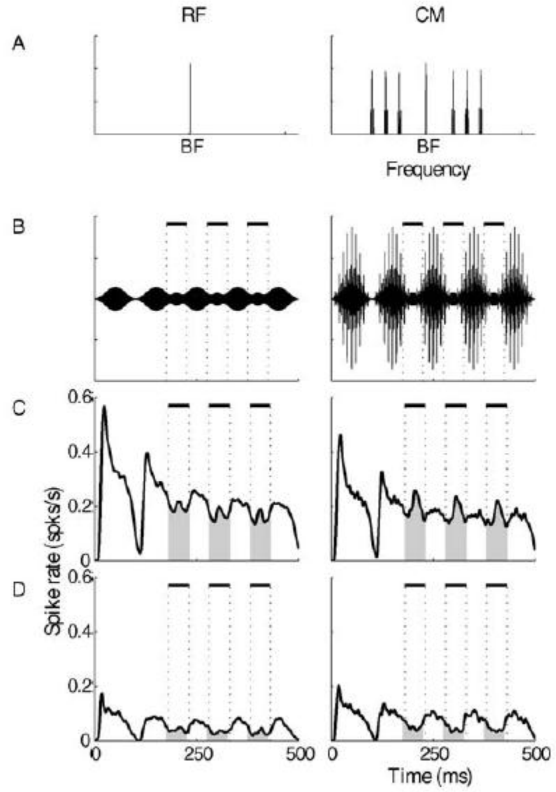


Figure 3: (Figure taken from [14]) Responses of chopper (C) and onset neurons (D) on the stimuli shown in (B) with the spectra presented in (A) for the reference condition (RF) and the comodulated condition (CM) where the masker consists of amplitude modulated tones. The signal are three 50 ms tone pips positioned in the minima of the masker indicated by horizontal black bars.

unit. They receive excitatory input from their best frequency and inhibitory input from the onset units abbreviated as WBI for wide band inhibitor. The wide band inhibitor receives input from a broad frequency range.

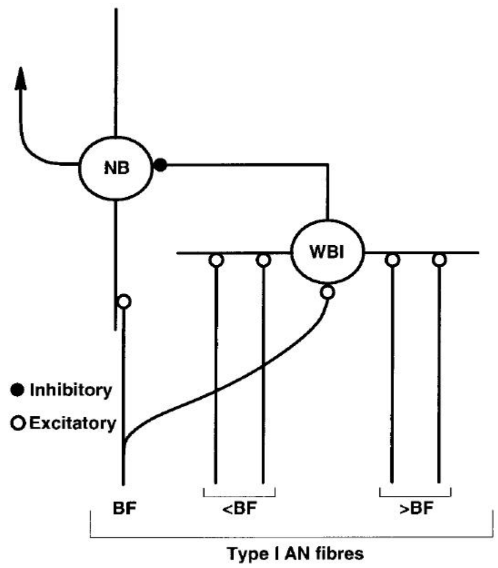
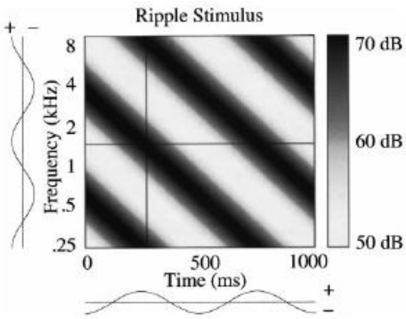


Figure 4: (Figure taken from [13]) Model to realize wideband inhibition. Narrowband unit (NB) receives excitatory input from their best frequency and inhibitory input from the onset units abbreviated as WBI for wide band inhibitor. The WBI receives input from a broad range of frequencies.

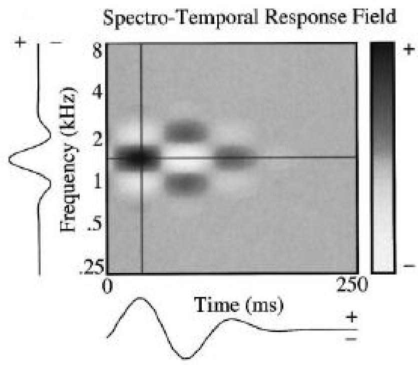
The effect of the combination for these two cells depends on the noise statistics. For the reference condition only a narrowband stimulus around the best frequency of the narrow-band unit is present. The response of the wide band inhibitor to narrow band noise is small. This results in a small inhibitory effect. Uncorrelated flanking bands have their maxima at different positions. The flanking bands and the on-frequency masker stimulate the wide band inhibitor continuously. Since inhibition is continuous, the signal and the noise masker are inhibited in equal measure. In the comodulated condition the maxima and the minima of all noise bands are temporally aligned. The response of the wide band inhibitor reflects the maskers structure. The response and with that the inhibitory effect is high during times with high masker energy and low if the masker energy is low. This reduces the masker effectiveness, but has little effect on the signal. The threshold for the signal masked by comodulated noise bands decreases.

2.4. Sweep stimuli

Most CMR experiments until now have studied hearing thresholds using time independent spectra. The goal of this experiment is to investigate CMR with time varying spectra. The question is whether CMR is still present if signals are used, which cover a range of frequencies. One possibility for a signal to cover several auditory filters in a short time is a sweep.



(a) Moving spectral ripples with a ripple velocity of 2Hz and a ripple density of 0.4 cycles per octave with a 10dB amplitude modulation.



(b) Idealized spectro-temporal response function of a single neuron

Figure 5: (Both figures from [3]) Example of a ripple stimulus and an idealized spectro-temporal response function.

One realization of sweeped stimuli are so called ripple stimuli or moving ripples. Ripple stimuli are sounds with sinusoidal spectral envelopes on a logarithmic frequency axis that are moving upwards or downwards with a constant velocity. Parameters characterizing a ripple are ripple density and ripple velocity. Ripple density is a measure for the spacing of ripples (minima and maxima) on the frequency axis and ripple velocity indicates the temporal modulation. A downward moving ripple stimulus is shown in figure 5a.

According to Chi et al [2] any arbitrary signal with a rich spectro-temporal envelope can be decomposed to a sum of ripple stimuli. The response of a single neuron to such a spectro temporally rich stimulus can be predicted by linear superposition of the responses to ripple stimuli with varying ripple densities and velocities. In human speech not only temporal, but also spectral modulations carry information and influence the speech intelligibility [2]. Ripple stimuli combine both types of modulation and therefore play an important role in investigating and modelling speech intelligibility. Figure 5b shows an idealized spectro-temporal response function (STRF) of a single neuron. An STRF characterizes the response of a neuron when stimulated with an arbitrary signal.

If the hypothetical neuron presented in figure 5b is stimulated with a tone at for example 1.5 kHz it will fire at the beginning of the signal. After 50 ms it will stop firing and is inhibited due to the previous activity. The inhibitory region causes a higher barrier potential, such that it is more difficult to excite the neuron. By designing a moving ripple stimulus with correct ripple density and ripple velocity these inhibitory regions could be avoided. Didier et al [3] found that neurons on the auditory path of ferrets show the highest firing rates when stimulated with ripple stimuli.

2.5. Grouping

One important mechanism that can be seen as an explanation for CMR is object grouping. The brain is able to combine not only visual objects, such as two parallel lines, but also auditory objects which show the same characteristic attributes. Important cues for grouping auditory signals are for example the starting time, amplitude modulation and a similar destination. Since amplitude modulation is an important cue for grouping, CMR effects are difficult to detach from grouping effects. Grose and Hall [7] have tried to separate the two effects, by introducing paradigms which prevent grouping. They seggregated the masker into two separate streams. One with the on-frequency masker including the signal and another containing the flanking bands. They found that CMR can be completely suppressed, when the starting and ending time of the on-frequency masker and the flanking bands differ more than 50 ms. This suggests that CMR cannot be seen as an individual effect, but is a part of grouping. According to Grose et al [6] harmonicity and frequency modulation have little or no effect on grouping. Since amplitude modulation is a cue for grouping, the brain combines the different masking bands to one auditory object and is able to detach it from the signal. Therefore CMR experiments are always also grouping experiments. CMR experiments with sweeped stimuli do not alter this. They can be designed to have the main attributes, that are important for grouping such as having the same destination, starting at the same time and containing a time-coherent amplitude modulation.

3. Experiment 1

3.1. Procedure

This experiment investigates the effect of sweeped stimuli on CMR by varying the sweep velocity of signal and masking bands. Four normal hearing subjects between the age of 20 and 27 participated in the experiment. All had previous experience with psychoacoustical measurements and at least 2 hours of training with similar stimuli. A three interval alternative forced choice 1up-2down procedure was used to estimate thresholds. The measured thresholds corresponded to 71 percent correct answers [9]. Four runs were recorded for each of the subjects. A run contained a set of all 22 stimuli used in both experiments. The mean thresholds of the six last reversal points per measurement were used to estimate the threshold. One trial consisted of three intervals, each of 500 ms duration, one containing the signal+masker and the other two only the masker. Panel lights indicated each interval. The participant was asked to specify the interval including the signal, by typing the number of the interval. A feedback (correct/false) was given after each answer via the monitor. The sound pressure level of the signal started at 80 dB, which was 20 dB higher than the sound pressure level of each noise band. It was lowered or raised by 4 dB until the first turning point, allowing a fast approach to the participants hearing threshold. After each turning point the stepsize was reduced. The actual recording of the threshold took place with a stepsize of 1 dB.

The experiment was carried out in a double walled sound proof chamber equipped with a monitor, a keyboard and plug for headphones, which were connected to a computer outside. For sound presentation a RME DIGI96/8 PAD soundcard was used. The stimuli were presented via a Sennheiser HD650 after amplification (-15 dB) with a Tucker and Davis HB7. The set-up was calibrated with a Brüel&Kjær No.4134 measurement microphone connected to display equipment Brüel&Kjær 2610. For calibration the headphones were adjusted to a Brüel&Kjær 2615 coupler and a 1 kHz sinusoid was generated and displayed with the headphones. The measured sound pressure level (SPL) was used to adjust the level of the stimuli. The starting level of the signal was set to 80 dB SPL and each of the masker bands to 60 dB SPL. The level of the masker including the four flanking bands in addition to the on-frequency masker was 67 dB.

3.2. Stimuli

Throughout both experiments sounds were generated digitally in Matlab with a 44100 Hz sampling rate. They were presented in intervals of 500 ms including a 50 ms rise and fall time, realized by multiplication with a Hann-window. The signal, which had to be detected was masked by one or several narrow noise bands. Each noise band was multiplied noise. A low pass noise band was generated and, depending on the condition, was multiplied with a sweep or sinusoid. This was equivalent to a convolution in the frequency domain and transformed the low pass noise onto a band pass noise with the center frequency of the sinusoid or sweep. The low pass noise was generated by assigning random amplitudes and phases drawn from a uniform distribution to frequency bins between 2 and 12 Hz. The frequency resolution for a 500 ms signal with a sampling rate of 44100 Hz is 2 Hz. The bin for the 0 Hz component was set to zero, in order to prevent superposition effects between the carrier frequency and the 0 Hz component of the noise. The outcome was a bandpass noise with a very narrow band gap (4 Hz). After multiplication with a sinusoid or sweep the result was a 24 Hz wide noise band around the frequency of the sinusoid or sweep. The sweep was chosen to be a logarithmic chirp, which means the frequency increased exponentially over time. The range of frequencies covered was a linear function of time, when plotted on with logarithmic frequency axis. The instantaneous frequency for the logarithmic sweep is calculated in equation 1. This instantaneous frequency is used as a frequency in a cosine function including a random phase to generate the sweep.

$$f(t) = f_{begin} \cdot \left(\frac{f_{end}}{f_{begin}} \right)^{\frac{t}{t_{end}}} \tag{1}$$

To adjust the levels of the masking bands, the amplitude of each noise sample was normalized by dividing by the mean amplitude of the 500 ms noise band and multiplication with the desired signal level. The level of each masking band was set to 60 dB SPL. The masker for the uncorrelated and the comodulated situations, which contained five masking bands was accordingly 67 dB. The conditions for both experiments were mixed and were presented in a random order during each run.

In the first experiment three different masker conditions were combined with six sweep velocities, resulting in 18 stimulus types. The spectrograms for these stimuli are shown in figure 6. Sweep velocities which are compared, were 0, +1/3, +1, +1.5, +2 and +2.5 octaves per 500 ms. In case of a sweep velocity of 0 octaves per 500 ms the signal which had to be detected was a 1 kHz sinusoid with a random phase. This condition was the classical CMR paradigm and was included to be able to relate the individual results to previous measurements discussed in literature as well as being able to compare the sweeped results with a regular CMR paradigm. For sweep velocities larger than zero the signal, which should be detected was a sweep centered around 1 kHz.

The sweep started below the signal center frequency and ended above it. The starting frequencies of the sweeps were shifted such that they reached center frequency after 250ms. In case for the signal and the on-frequency masker the center frequency was 1 kHz. The start frequency for each masking band was calculated as following:

$$f_{begin} = exp\left(log(f) - m \cdot \frac{t}{2}\right) \qquad (2)$$

where m was the gradient of the sweep, which is directly obtained from the sweep velocity, of the unshifted signal and f the desired center-frequency of 1 kHz and t the interval length in seconds. Start and end frequency for each noise band depend on the sweep velocity as well as their center frequency. After calculating the start frequency of a masking band the end frequency was shifted about the same $Deltaf = f - f_{begin}$. This ensured that the sweep velocity of 1/3 octaves was based on the original center-frequency. A composition of the frequency shifts for the different sweep velocities is given in table 1.

	0 octave $f_{start} - f_{end}/Hz$	**1/3 octaves** $f_{start} - f_{end}/Hz$	**1 octaves** $f_{start} - f_{end}/Hz$	**1.5 octaves** $f_{start} - f_{end}/Hz$	**2 octaves** $f_{start} - f_{end}/Hz$
OFM (1000)	1000-1000	891-1123	707-1414	595-1682	500-2000
FB1 (400)	400-400	365-449	282-566	238-673	200-800
FB2 (600)	600-600	535-674	324-849	357-1009	300-1200
FB3 (1400)	1400-1400	1247-1571	990-1980	832-2354	700-2800
FB4 (1600)	1600-1600	1425-1796	1131-2262	951-2691	800-3200

Table 1: Starting and ending frequencies for different sweep velocities for each masking band. The assigned center frequency of each noise band was reached after 250ms

In all conditions the signal is masked by a 24 Hz wide noise band centered around the signal frequency called on-frequency masker (OFM). The reference condition (RF) contained only the signal and the on-frequency masker at a frequency of 1000 Hz without flanking bands. For the other conditions four additional flanking bands were added. The flanking bands were distributed

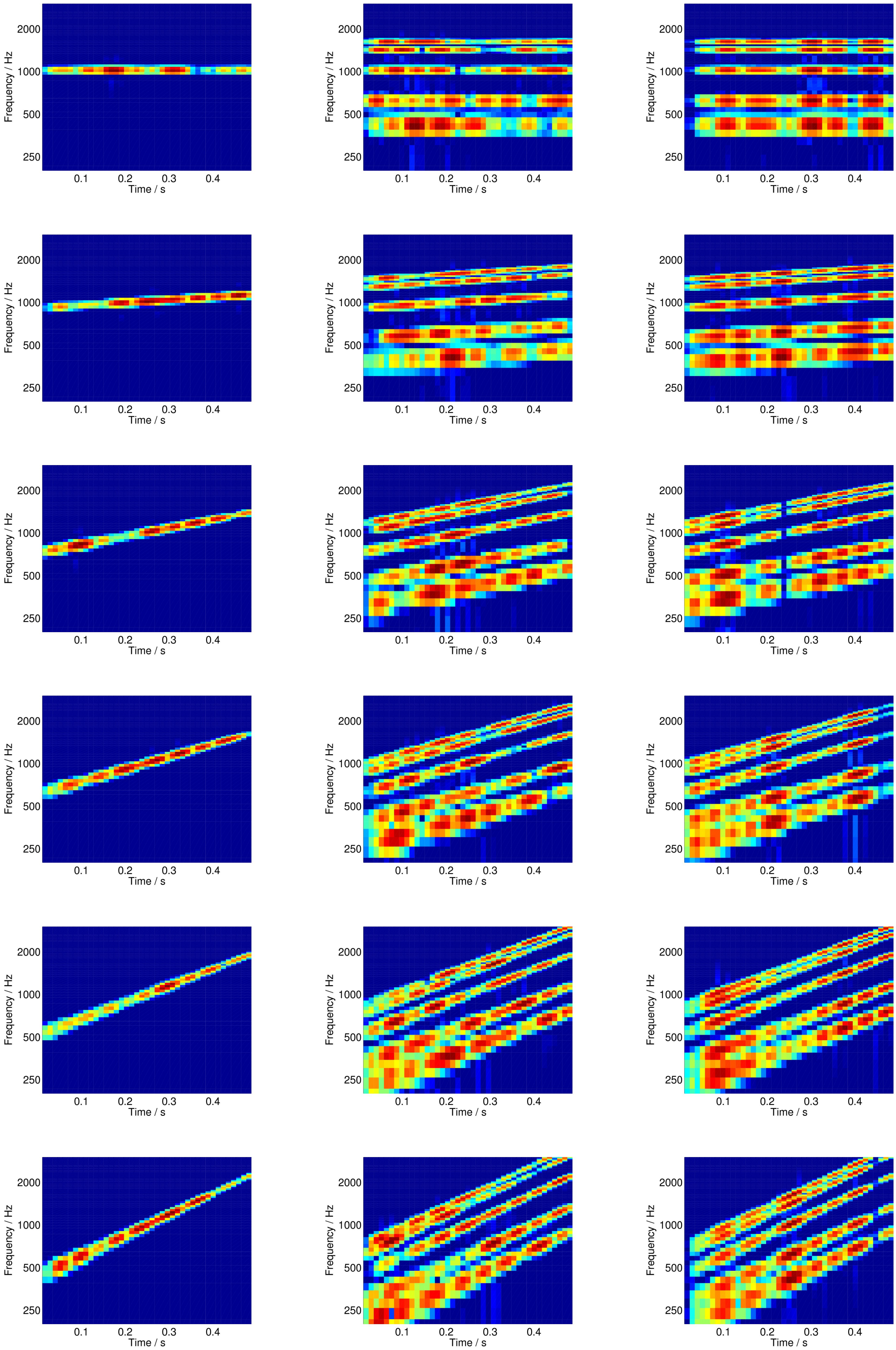


Figure 6: Spectrograms of the stimuli of experiment 1. Left column: RF condition, middle column: UN condition and right column: CM condition. The rows show the stimuli of the used sweep velocities (top down): 0, 1/3, 1, 1.5, 2, 2.5 octaves per interval

symmetrically above and below the on-frequency masker at center frequencies of 400 Hz, 600 Hz, 1400 Hz and 1800 Hz. In the uncorrelated condition (UN) a low pass noise was created for each of the flanking bands individually resulting in five uncorrelated noise bands. For the comodulated situation (CM) the same low pass noise was multiplied with sweeps at the various center frequencies resulting in a masker with five comodulated noise bands.

3.3. Results and Discussion

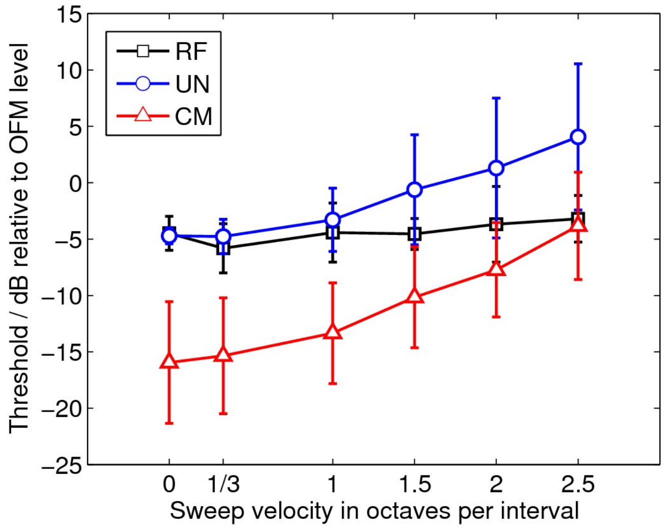


Figure 7: Mean thresholds for all subjects for the RF, the UN and the CM situation depending on the sweep velocity. Thresholds are in dB relative to the level of 60 dB of the on frequency masker. The error bars indicate the standard deviation from the mean of all subjects after averaging all trials of each

The mean thresholds for all subjects for the reference (RF), the uncorrelated (UN) and the comodulated (CM) situation depending on the sweep velocity are shown in figure 7. Thresholds in dB are relative to the level of the on frequency masker (OFM), which was 60dB SPL. The error bars indicate the standard deviation from the mean.

The reference threshold stays fairly constant over the different sweep velocities at -4 dB relative to the OFM level. The uncorrelated as well as the comodulated thresholds increase with increasing ripple velocities. The uncorrelated threshold increases from -4 dB for 0 sweep velocity to +4 dB for a sweep velocity of 2.5 octaves per intervall. The threshold for the comodulated situation increases from -16 dB to -5 dB. The comodulation masking release CMR_{RF-CM} and CMR_{UN-CM} is approximately 10 dB for the classical CMR experiment with sweep rate of 0 octaves per intervall.

The CMR_{CM-UN} stays constant for all sweep velocities used. CMR_{RF-CM} constantly decreases. It is zero for a ripple velocity of 2.5 octaves. The standard deviations of up to ±6 dB especially for the comodulated situation and the uncorrelated condition for higher sweep velocities can be explained by inter individual discrepancies. The individual data are shown in figure 8.

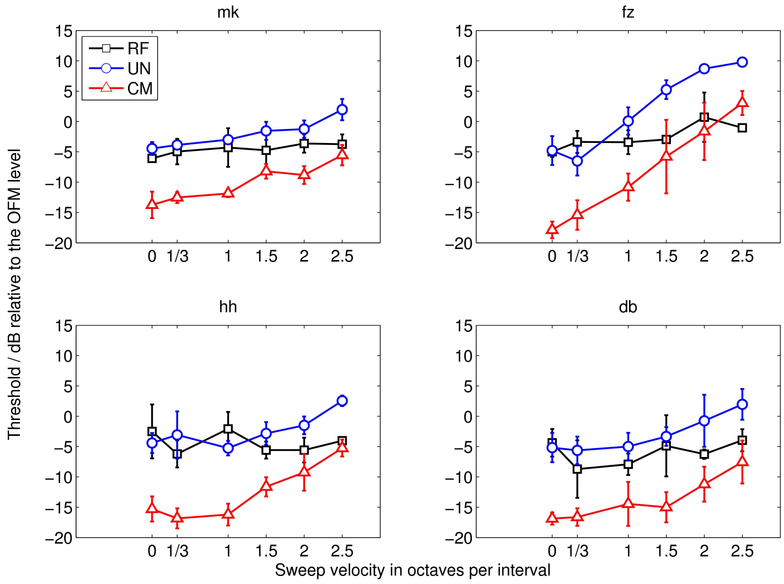
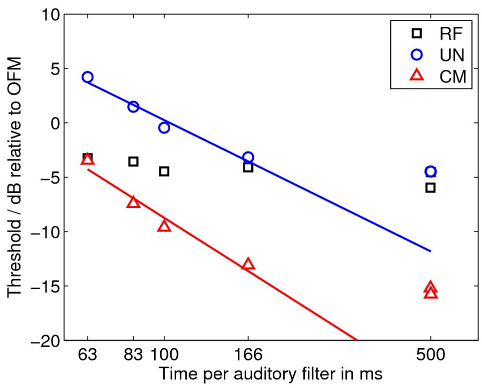


Figure 8: Individual thresholds for all subject for the RF, the UN and the CM situation in dependence of the sweep velocity. Thresholds are in dB relative to the level of 60 dB of the on frequency masker. The error bars indicate the standard deviation from the mean over all trials.

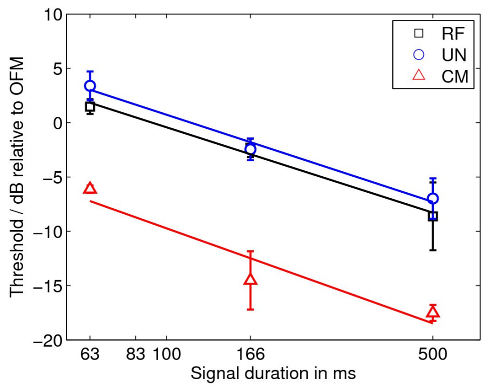
For all subjects, the uncorrelated and the reference threshold are very close to each other at a sweep velocity of zero. The comodulated threshold is considerably lower. The CMR_{RF-CM} for a sweep velocity of zero is -14 dB for three subjects and -8dB for one subject. All subjects show increasing uncorrelated and comodulated thresholds with increasing sweep velocity. The inter individual differences concerning the rate of increase are large. Subject fz shows the strongest increase of more than 20 dB for the comodulated threshold between zero sweep velocity and 2.5 octaves per interval. The other subjects show increases between 7 and 10 dB. For three subjects, a small CMR_{RF-CM} is still present for a sweep velocity of 2.5 octaves per interval. The comodulated threshold for the fourth subject is higher than the reference threshold resulting in a negative CMR_{RF-CM}.

A classical CMR paradigm corresponds to a ripple velocity of zero. 10 dB CMR_{RF-CM} is a similar result as for example found by Schooneveldt and Moore [12] in a flanking band experiment for a flanking band close to the on-frequency masker. The reference threshold they found for 25 Hz wide noisebands was -4 dB relative to the on-frequency masker, which is consistent with the reference threshold here. Ernst and Verhey [4] found a different relative reference threshold of almost +4 dB. Differences to the data of Ernst and Verhey might be due to a slightly different bandwidth (20 Hz) and a much lower overall level of the on-frequency masker (20 dB SPL).

With increasing sweep velocities the time for which a signal stays within one auditory filter is reduced. McFadden studied the effect of signal duration on CMR in [11]. He found that thresholds for the uncorrelated and comodulated situation increase with decreasing signal duration with a slope of 9.3 dB per decade for signal durations between 75 and 375 ms. Assuming rectangular auditory filters with 1/3 octave bandwidth, the duration of the signal within one auditory filter in the present experiment varies from 500 ms for sweep velocities of 0 and 1/3 to 63ms per auditory filter for a sweep velocity of 2.5 octaves. The thresholds from figure 7 are replotted now as a function of time per auditory filter. These data are shown in figure 9a.



(a) Thresholds from figure 7 replotted as a function of time per auditory filter assuming rectangular filters with 1/3 octave bandwidth. The lines are best fitting lines through the data points excluding 500 ms.



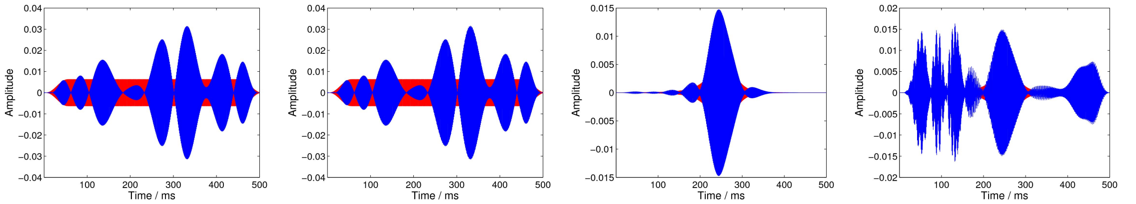
(b) Measured thresholds (of two subjects) for a sweep velocity of zero, but varying signal duration.

Figure 9: Thresholds dependencies of time per auditory filter and signal duration. Thresholds in dB are expressed relative to the level of the on-frequency masker.

The 500 ms point actually combines two of the measured conditions. For a sweep velocity of 0 and 1/3 octaves per interval, the signal does not leave the auditory filter centered at 1000 Hz. The thresholds for these two sweep velocities are very similar supporting the hypothesis that the time per auditory filter determines the threshold. The straight line is a best linear fit through the measured data on a logarithmic time scale. The 500 ms point is excluded, because an integrating effect by a threshold reduction can no longer be observed and a saturation effect begins. This corresponds to McFadden's analysis, who also excluded 600ms signal duration from the calculation of the best fit slope. The slope of of the best fitting line through the data of this experiment are 17.5 dB per decade for the uncorrelated situation and 22.3 dB per decade for the comodulated situation. They are steeper than those found by McFadden, but in tendency the reduced time per auditory filter could serve as an explanation for increased thresholds of the uncorrelated and the comodulated situation. The reference condition is unchanged, which is contradictory to the assumption that the time per filter is the reason for an increase of thresholds for the uncorrelated and comodulated situation. To verify the hypothesis that time per auditory filter is an explanation at least for the uncorrelated and the comodulated condition an additional small experiment was carried out with two of the four subjects.

In this sub-experiment the sweep velocity was kept constant at zero, but instead the signal duration was varied. The signal was positioned temporally in the middle of the 500 ms noise. The signal durations used were 63 ms corresponding to the time per auditory filter for a sweep velocity of 2.5 octaves per interval, 166 ms corresponding to 1 octave per interval and 500 ms corresponding to sweep velocities of 0 and 1/3 octavs per interval. The rise and fall times of the Hann window were changed to 10 ms. All other parameters were the same as described in section 3.2. The averaged results of this experiment are shown in figure 9b. The thresholds of all conditions increase with decreasing signal duration with a similar slope. Increased thresholds for the uncorrelated and comodulated condition in the experiment varying the sweep velocity show the same tendencies as the thresholds for the duration experiment. Thresholds of the reference condition are not replicated. For the duration experiment they decrease with a similar slope as the uncorrelated and the comodulated condition. In order to be able to explain a constant reference threshold for all sweep velocities one has to conclude that more information can be used for the detectability of the signal, than is present within one auditory filter. Therefore multiple looks into consecutive filters are needed to explain this constant reference threshold. The principle of multiple looks instead of an integrating effect over time is discussed by Buus [1]. He found that temporal integration can be described by optimal combinig multiple looks onto brief signal segments. The constant RF slope supports this and even suggests that each look can be into a different auditory filter.

Obviously this advantage can not be used for the uncorrelated and the comodulated condition. In order to investigate this difference between the reference on the one hand and the comodualted and the uncorrelated conditions on the other hand, the signals were sent through a fourth order gamma tone filter centered at 1 kHz. The filter responses to sweep velocities of 0 and 2 ocatves per interval for the reference and the comodulated condition are shown in figure 10. The signal



(a) RF 0 Octaves (b) CM 0 Octaves (c) RF 2 Octaves (d) CM 2 Octaves

Figure 10: Filter response to sweeped stimuli for 0 and 2 ocatves per interval for the RF and the CM condition including the signal in red

is drawn in red partly hidden by the noise masker drawn in blue. Obviously there is a difference between stationary signals and sweeped signals for the conditions including flanking bands. For sweeped stimuli energy from the flanking bands is put into the auditory filter centered around the signal frequency. While the signal sweeps from below the center frequency of the filter to above, some of the flanking bands reach this filter before or after the signal. The amount of energy, which is deposited in the filter depends on the magnitude of the sweep velocity and the distance between the flanking bands and the on-frequency masker. Considering this post-masking effects of the flanking bands for sweeped condition might play a role. Another interpretation is temporal uncertainty. For the reference condition only one auditory filter is activated. The subjects know in which filter to look for the signal. For conditions including flanking bands it is not clear which of the activated filters includes the signal.

Apparently humans have the ability to look into several auditory filters and integrate the information from all of them. Otherwise it is not possible to explain a constant reference threshold for the sweeping conditions, but an increased threshold if the signal duration is varied. With respect to the hypothesis of wide band inhibition as a physiological model to explain CMR this means that several narrowband units with different best frequencies are involved in the encoding of the stimuli. All these narrowband units contribute to the detectability of the signal. The information has to be collected on a higher stage and is then analyzed. The wide band inhibitor should not be influenced by the sweeping velocity of the signal, because it is sensitive to broad band signals.

4. Experiment 2

4.1. Procedure

The same subjects as in experiment 1 participated in this experiment. The presentation of the stimuli for this experiment was randomly mixed with the stimuli of the previous experiment and measured simultaneously. Setup and calibration were as described in section 3.1.

4.2. Stimuli

The noise bands were generated as in experiment 1. In the second experiment the sweep velocity was kept constant at 1/3 octaves per interval. As before each flanking band was shifted in frequency as calculated in equation 2. Only one low pass noise was generated to create the flanking bands. This corresponds to comodulated flanking bands. Before transforming the low pass noise to the correct frequency by multiplying it with a sweep, each flanking band was shifted in time. The amount of time shift for each flanking band was chosen such that the alignment of maxima and minima is not completely lost. Instead of a temporal alignment for the comodulated condition indicated by vertical lines across the stimulus spectrogam, these lines are tilted. The temporal comodulation is transformed into a spectro-temporal modulation comparable to the spectro-temporal modulations of ripple stimuli. The slope of the tilted line is a measure for the amount of timeshifts. In the classical comodulated situation (CM) minima and maxima of the entire frequency range are temporally aligned and can be connected by vertical lines. Therefore, using this description of the spectro-temporal modulations (STM) the regular comodulated condition can be described by a spectro-temporal modulation with a slope of infinite octaves per intervall. Other slopes created by shifting the flanking bands were: 1.5, 3, 14 and 26 octaves per interval. For a 1.5 octaves per interval slope the time shifts for each flanking band are calculated as follows:

$$\Delta t = \frac{log(f_{signal}) - log(f)}{m + 1/m} \tag{3}$$

where f_{signal} is the signal center frequency, f is the center frequency of the flanking band that is shifted in time and m denotes the gradient of the sweep as described above. The Δt calculated here is in seconds and has to be converted to samples by multiplication with the sampling frequency. The time shift has to be calculated individually for each flanking band. Conditions with steeper slopes were created by dividing the timeshifts calculated in equation 3 by several factors. Deviding the timeshifts Δt of each flanking band by 2 yields a slope of 3 octaves per intervall, a division by 10 corresponds to 14 octaves per interval and a slope of 26 octaves per interval is created by dividing the timeshifts by 20. The time shifts for each flanking band is given in table 2. Spectrograms for these stimuli are shown in figure 11.

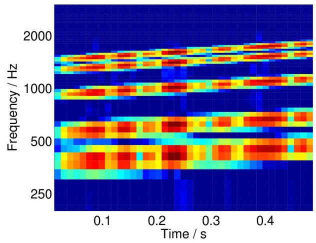
	-1.5: Δt	**-3:** Δt	**-14:** Δt	**-26:** Δt	**CM:** Δt
OFM (1000)	0 ms	0 ms	0 ms	0 ms	0 ms
FB1 (400)	349 ms	174 ms	35 ms	17 ms	0 ms
FB2 (600)	194 ms	97 ms	19 ms	10 ms	0 ms
FB3 (1400)	-128 ms	-64 ms	-13 ms	-6 ms	0 ms
FB4 (1600)	-179 ms	-89 ms	-18 ms	-9 ms	0 ms

Table 2: Time shifts for different sweep velocities for each masking band. The different rows show time shifts with slopes of -1.5, -3 -14, -26 octaves per interval

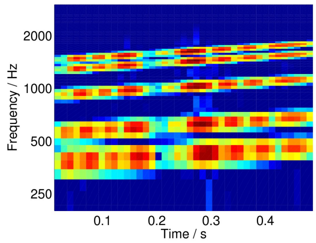
For a sweep up the flanking bands above the signal frequency have to be shifted backwards and the flanking bands below the signal frequency have to be shifted forward. In order to be able to shift the noise bands in time, the low pass noise for the spectro-temporal modulated condition was 2 seconds long and for the on-frequency masker a 500 ms long portion is cut out. The longer noise duration results in a frequency resolution of 0.5 Hz. As in in the first experiment the bandwidth of the low pass noise is limited from 2 to 12 Hz. After the band pass noise is shifted in time, according to the different conditions, it is multiplied with the sweep to position the noise band at the correct center frequency.

4.3. Results and Discussion

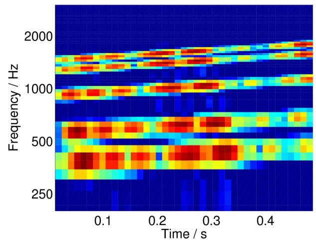
The mean thresholds for all subjects in dependence of the ripple slope of the inclination are shown in figure 12. The threshold in dB is the relative to the threshold of the 60 dB SPL of the sound pressure level for the on-frequency masker. The error bars indicate the standard deviation from the mean over all subjects after averaging all trials for each subject. The thresholds increase with decreasing slope. All subjects show a reduction of CMR_{UN-CM} with an increase of the spectro-temporal tilted modulation. The threshold for the slope of -1.5 octaves per interval and -3 octaves per interval condition is the same or slightly higher as for a uncorrelated situation and therefore there is no CMR_{UN-CM} at all for these conditions.



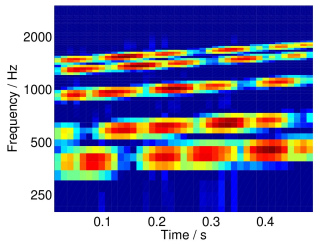
(a) slope: $-\infty$ octaves per intervall



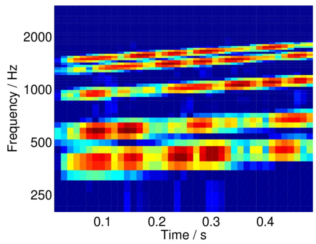
(b) -26 octaves per interval



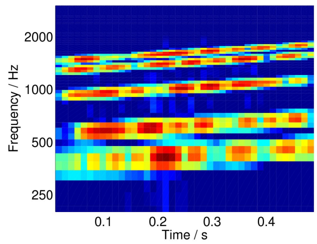
(c) -14 octaves per intervall



(d) -3 octaves per interval



(e) -1.5 octaves per intervall



(f) UN

Figure 11: Spectrograms of the ripple like stimuli with different slopes used in experiment 2.

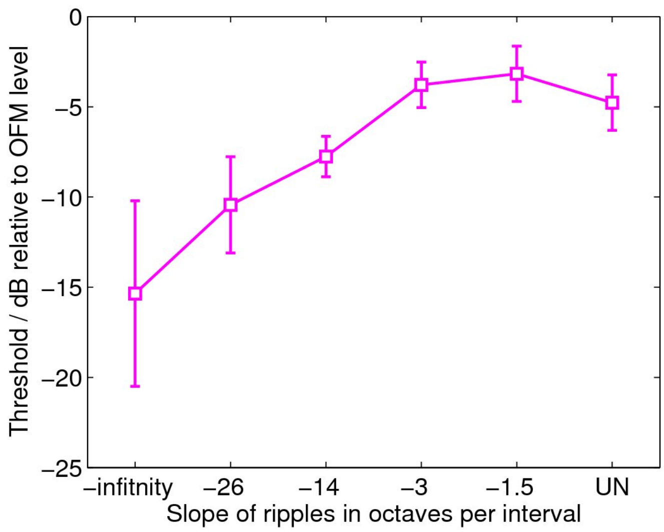


Figure 12: Mean thresholds for all subjects in dependence of the slope on the tilt. Thresholds in dB are expressed relative to the level of the on-frequency masker. The error bars indicate the standard deviation from the mean over all subjects after averaging all trials for each subject

Changing the slope of the comodulation is realized by introducing time shifts to each flanking band. These timeshifts seem to destroy the masking release obtained in a comodulated situation. Thus spectro-temporal modulations can not be used to achieve a better signal detectability. Looking at table 2, it is not possible to assign a single value of timeshift to each condition. In general however the time shifts become longer for stronger delayed flanking bands. In order to compare these findings to CMR experiments with time delayed flanking bands in the literature an estimate over all time shifts is needed. One simple estimate is to use the time shift of the flanking band with the smallest delay and and ignore the effect of the other more delayed flanking bands. Larger delays lead to higher thresholds. Moore and Schooneveldt analyzed the effect of a time delayed flanking band using one flanking band [12]. They combined several bandwidths with several center frequencies of the flanking band and found that for 25 Hz bandwidth CMR decreases from 3.5 dB for 0 delay to 1.5 dB for a 20 ms delay between the on-frequency masker and the flanking band. They argued that the reduced correlation of the noise bands, depending on bandwidth and time shift is the reason for the reduced CMR. Considering these results the increase in thresholds for flat slopes in the present study can be explained by increasing decorrelation between the on-frequency masker and the closest flanking band.

5. Summary and Conclusion

Experiment 1 has shown that CMR occurs with sweeped stimuli. The reference threshold is not influenced by sweep velocity whereas uncorrelated and comodulated thresholds increase. From this follows that CMR_{RF-CM} is reduced and is zero for a sweep velocity of 2.5 octaves per interval. CMR_{UN-CM} is constant over all sweep velocities. These results can be explained by assuming the ability to integrate information out of different auditory filters. The time for which the signal stays within one auditory filter is shorter the higher the sweep rate. The shorter time can be compensated by consecutively adding the information from all filters. However postmasking effects of flanking bands sweeping through the filters or temporal uncertainty of where to look for the signal have to be taken into consideration in order to explain increasing uncorrelated and comodulated thresholds. To verify this explanation further experiments with non- sweeped signals, but including postmasking paradigms is necessary. Also the hypothesis of temporal uncertainty has to be verified with follow up experiments in which an uncertainty is introduced. Such a temporal uncertainty could be introduced by keeping constant masking bands and making the signal jump between the flanking bands. Also not considered in this experiment were sweep down stimuli.

Experiment 2 was set up to investigate the effect of spectro temporal modulation. By tilting the lines connecting temporal alligned modulations thresholds are increased relative to the comodulated condition. This effect can be put down to an increasing time delay between the on frequency masker and the flanking bands. This explanation has to be verified by further experiments using different bandwidths for the noise bands. Wider bandwidths should result in a stronger increase in thresholds for the same time shifts. Spectro-temporal modulations, similar to ripple stimuli, do not enhance signal detectability.

In summary are all these findings in agreement with the principle of wideband inhibition as a neural basis for CMR. But the model introduced in chapter 2.3 has to be complemented with a higher stage that collects the information out of several narrowband units and is able to combine them.

A. Appendix

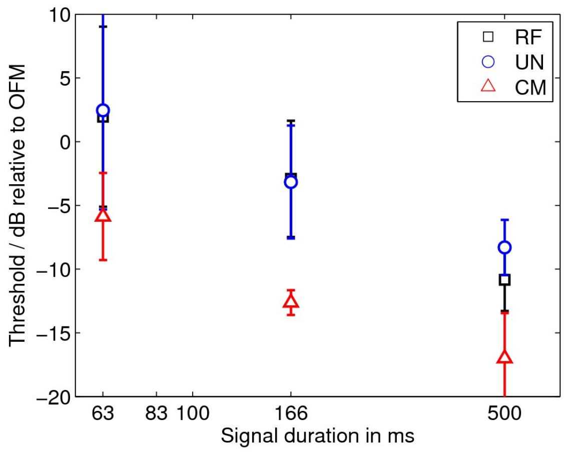


Figure 13: Thresholds for subject mk in the duration experiment as described in the subexperiment for experiment 2. Threshold in dependence of duration of the signal. Thresholds in dB are expressed relative to the level of the on-frequency masker. The error bars indicate the standard deviation from the mean over all trials.

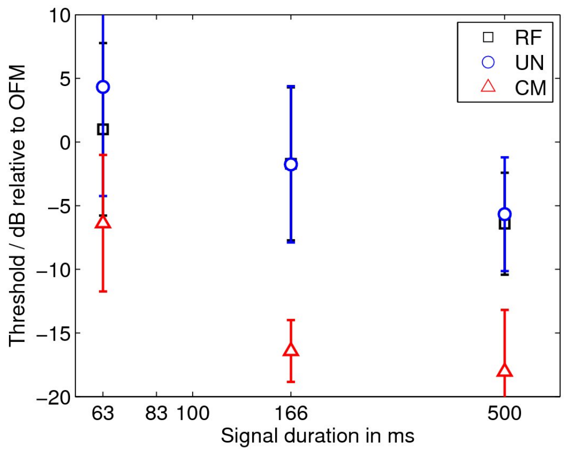


Figure 14: Thresholds for subject hh in the duration experiment as described in the subexperiment for experiment 2. Threshold in dependence of duration of the signal. Thresholds in dB are expressed relative to the level of the on-frequency masker. The error bars indicate the standard deviation from the mean over all trials.

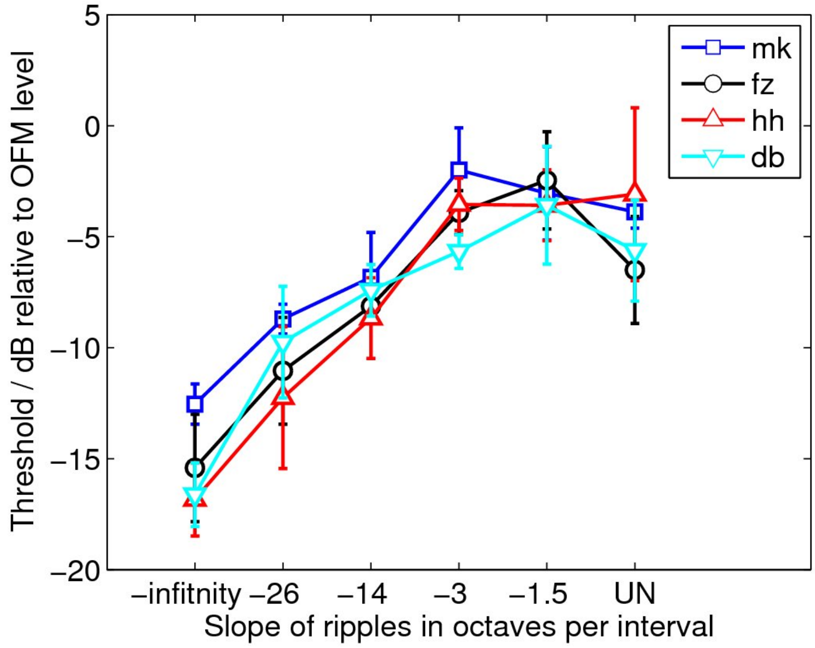


Figure 15: Individual thresholds for all subjects in dependence of the slope on the tilt as described in the second experiment. Thresholds in dB are expressed relative to the level of the on-frequency masker. The error bars indicate the standard deviation from the mean over all trials.

References

[1] S. Buus. Temporal integration and multiple looks, revisited: Weights as a function of time. *J. Acoust. Soc. Am.*, 105:2466–2457, 1999.

[2] T. Chi, Y. Gao, M. C. Guyton, P. Ru, and S. Shamma. Spectro-temporal modulation transfer functions and speech intelligibility. *J. Acoust. Soc. Am.*, 106:2719–2732, 1999.

[3] D. A. Depireux, J. Z. Simon, D. J. Klein, and S. A. Shamma. Spectro-temporal response filed characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol*, 85:1220–1234, 2001.

[4] S. M. A. Ernst and J. Verhey. Role of suppression and retro-cochlear processes in comodulation masking release. *J. Acoust. Soc. Am.*, 120:3843–3852, 2006.

[5] H. Fastl and E. Zwicker. *Psychoacoustics*. Springer, Heidelberg, third edition, 2007.

[6] J. H. Grose, J. W. Hall, and L. Mendoza. Percepual organization in a comodulation masking release interference paradigm: Exploring the role of amplitude modulation, frequency modulation, and harmonicity. *J. Acoust. Soc. Am.*, 97:3064–3071, 1995.

[7] J.H. Grose and J.W. Hall. Comodulation masking release: Is comodulation sufficient? *J. Acoust. Soc. Am.*, 93:2896–2902, 1993.

[8] W. M. Hartmann. *Signals, Sound, and Sensation*. Springer, New York, fifth edition, 2005.

[9] H. Levitt. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49:467–477, 1971.

[10] J. L.Verhey, T. Dau, and B. Kollmeier. Within-channel cues in comodulation masking release (cmr): Experiments and model predictions using a modulation-filterbank model. *Acoustical Society of America Journal*, 106:2733–2745, November 1999.

[11] D. McFadden. Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band. *J. Acoust. Soc. Am.*, 80:1658–1667, 1986.

[12] B. C. J. Moore and G.P. Schooneveldt. 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, 1990.

[13] D. Pressnitzer, R. Meddis, R Delahaye, and I. Winter. Physiological correlates of comodulation masking release in the mammalian ventral cochlear bucleus. *J. Neuroscience*, 21:6377–6386, 2001.

[14] J. L. Verhey, D. Pressnitzer, and I. M. Winter. The psychophysics and physiology of comodulation masking release. *Exp. Brain Res.*, 153:405–417, 2003.

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe.