

Simultaneous masking additivity for short Gaussian-shaped tones: Spectral effects^{a)}

Bernhard Laback,^{b)} Thibaud Necciari, and Peter Balazs

Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

Sophie Savel and Sølvi Ystad

Laboratoire de Mécanique et d'Acoustique, CNRS, UPR 7051, Aix-Marseille University, Centrale Marseille, F-13402 Marseille Cedex 20, France

(Received 13 September 2012; revised 13 June 2013; accepted 14 June 2013)

Laback *et al.* [(2011). *J. Acoust. Soc. Am.* **129**, 888–897] investigated the additivity of nonsimultaneous masking using short Gaussian-shaped tones as maskers and target. The present study involved Gaussian stimuli to measure the additivity of *simultaneous* masking for combinations of up to four spectrally separated maskers. According to most basilar membrane measurements, the maskers should be processed linearly at the characteristic frequency (CF) of the target. Assuming also compression of the target, all masker combinations should produce excess masking (exceeding linear additivity). The results for a pair of maskers flanking the target indeed showed excess masking. The amount of excess masking could be predicted by a model assuming summation of masker-evoked excitations in intensity units at the target CF and compression of the target, using compressive input/output functions derived from the nonsimultaneous masking study. However, the combinations of lower-frequency maskers showed much less excess masking than predicted by the model. This cannot easily be attributed to factors like off-frequency listening, combination tone perception, or between-masker suppression. It was better predicted, however, by assuming weighted intensity summation of masker excitations. The optimum weights for the lower-frequency maskers were smaller than one, consistent with partial masker compression as indicated by recent psychoacoustic data. © 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4812773]

PACS number(s): 43.66.Dc, 43.66.Ba [ELP]

Pages: 1160–1171

I. INTRODUCTION

In this study, the additivity of simultaneous masking was measured for stimuli that are well concentrated in both the time and frequency domains, namely short Gaussian-shaped tones (referred to as “Gaussians”). Studying masking additivity with such stimuli is relevant for two main reasons. First, well-concentrated stimuli can be more flexibly arranged in the time-frequency space than temporally or spectrally broad stimuli while essentially avoiding spectro-temporal overlap of the maskers and the target. Therefore, they are well-suited for studying masking additivity in various closely spaced time-frequency configurations. Data collected with such stimuli may provide insight into the time-frequency dependencies of masking and the properties of spectro-temporal processing of the auditory system at short time constants. For example, they may help to better understand how the individual spectral components of a plosive in speech contribute to its total masking effect. Second, well-concentrated stimuli can be considered as elementary sounds. This property arises from time-frequency analysis schemes widely used in audio signal processing, such as the Gabor or wavelet transforms (e.g., Gröchenig, 2001). By means of such transforms any real-world sound can be

decomposed into a set of functions or “atoms,” for example Gaussians. In order to better understand and predict mutual masking effects between the atoms of a real-world sound (see, for example, Balazs *et al.*, 2010), it is worthwhile to study basic masking effects with such well concentrated stimuli. Previous studies focused on the spread of time-frequency masking (Necciari *et al.*, 2012) and the additivity of nonsimultaneous masking (Laback *et al.*, 2011) for Gaussians. In the present study, we investigate the additivity of *simultaneous* masking.

Masking additivity refers to the properties according to which the amount of masking produced by one masker adds with the amount of masking produced by another masker. Assuming linear additivity of masking in units of intensity, the masked threshold of a target should be 3 dB higher in the presence of two equally effective maskers (i.e., each masker alone causes the same masked threshold) than in the presence of each masker alone. In fact, in certain configurations, combining two maskers can result in higher masked thresholds than according to linear additivity (e.g., Penner, 1980; Humes and Jesteadt, 1989; Oxenham and Moore, 1995; Plack *et al.*, 2006; Plack *et al.*, 2008). The difference between the linear prediction and the actually measured masked threshold is referred to as the amount of “excess masking.” The commonly accepted origin of excess masking for nonsimultaneous maskers, assuming that maskers are processed independently before their effects are combined linearly at some higher stage, is that the target is subjected to a compressive nonlinearity (Penner, 1980). The more compressive the nonlinearity, the

^{a)}Portions of this work were presented at the Acoustics'08 Meeting in Paris, 2008.

^{b)}Author to whom correspondence should be addressed. Electronic mail: Bernhard.Laback@oeaw.ac.at

more excess masking occurs. In Laback *et al.* (2011), the additivity of nonsimultaneous masking was measured for a Gaussian target and combinations of up to four equally effective Gaussian maskers. On average, excess masking amounted to 13.5 dB for masker pairs and 26.0 dB for four maskers. These values are considerably greater than those reported in previous studies involving similar target sensation levels (8 dB) but much longer masker durations. We attributed this difference to the medial olivocochlear reflex (MOCR). When activated, the MOCR controls the cochlear gain via efferent connections to the outer hair cells (e.g., Backus and Guinan, 2006). It reduces the cochlear amplifier gain, which can result in an increased slope of the basilar membrane (BM) input/output function in a certain level region (e.g., Russell and Murugasu, 1997). The MOCR has an onset delay¹ of about 25 ms (Backus and Guinan, 2006). This delay is shorter than the between-stimuli intervals (i.e., between the onsets of the first masker and the target) involved in previous studies using long maskers, for which the MOCR may have been elicited and thus have reduced the cochlear gain. Conversely, the MOCR delay is longer than the between-stimuli intervals in Laback *et al.* (2011). Assuming that our data were unaffected by any MOCR-based gain reduction, the target was likely subjected to greater compressive nonlinearity than in past studies. This could have led to more excess masking. Some indication for such an increase in the amount of estimated compression for short relative to long maskers was recently observed in some experimental conditions of Plack and Arifianto (2010) using the additivity-of-forward-masking technique and it was reported in a cross-study comparison by Laback *et al.* (2011). It should be noted that those effects we and others attributed to the MOCR could also be due to some other post-cochlear adaptation mechanism(s). However, the temporal and spectral properties of the MOCR have been extensively described in the literature, which enables discussion of masking results according to the spectro-temporal characteristics of the particular stimuli under consideration. Therefore, the discussion of the present results focuses on the possible implication of the MOCR.

In the present study, we focused on the additivity of *simultaneous* masking for Gaussians. The methods and conditions were similar to those used in Laback *et al.* (2011) except that the maskers were separated in frequency rather than in time. Because it is hard to conceive how simultaneous maskers could be independently processed at the characteristic frequency (CF) of the target, the above-described explanation for the additivity of *nonsimultaneous* masking cannot apply here. Instead, an explanation based on simultaneous, thus, joint processing of the maskers and the target has been proposed (Oxenham and Moore, 1995). This explanation assumes that the BM response is compressive for tones around the CF but linear for tones remote from the CF as suggested by a number of studies (Rhode, 1971; Robles *et al.*, 1986; Ruggero *et al.*, 1992; Ruggero *et al.*, 1997; Russell and Nilsen, 1997; Rhode and Recio, 2000). All those studies reported linear processing for frequencies below -25% relative to the CF. The majority of studies also reported linear processing for frequencies above $+30\%$ relative to the CF. Only Rhode and Recio (2000) reported partial compression up to $+60\%$ relative to the CF. Furthermore, it

should be noted that this survey of the physiological literature refers to studies focusing on an area close to the basal end of the cochlea. Although studies are very sparse, the work by Rhode and Cooper (1996) suggests that compression may be less frequency selective towards the apical end of the cochlea relative to the basal end. The explanation for simultaneous excess masking further assumes that the masked thresholds both in the single-masker and the combined-masker conditions are given by a fixed target-to-masker ratio in the cochlear output at the target CF. Then, one has to distinguish two cases in simultaneous masking additivity. In the first case, when maskers are close in frequency to the target, the jointly processed maskers undergo the same amount of compression as the target. Because the increase in masker-evoked excitation when adding a second (or more) masker(s) is compressed like the target, no excess masking can be expected and linear additivity (masked threshold of the target increases by 3 dB for two equally effective maskers) applies. In the second case, when maskers are sufficiently spectrally separated from the target, the response growth is compressive for the target but the responses to the maskers are assumed to be summed in intensity units at the target CF, which causes excess masking (Oxenham and Moore, 1995). In other words, while compression of the target is the basic requirement for nonlinear additivity to occur, the degree of excess masking depends critically on how the masker-evoked excitations are summed at the target CF. Intensity summation leads to maximum excess masking and weighted intensity summation with weights less than one (such as in a compressive system) results in reduced excess masking. Accordingly, previous studies reported excessive masking additivity for spectrally remote maskers and approximately linear additivity for spectrally close maskers (Zwicker and Herla, 1975; Moore, 1985; Humes *et al.*, 1992). We cite here only those studies that controlled potentially confounding factors such as off-frequency listening or combination-tone perception. Below, the outlined explanation for the excessive additivity of simultaneous masking for spectrally remote maskers is referred to as the “standard model of simultaneous masking.”

Another well-known effect related to the additivity of simultaneous masking is the so-called “upward spread of simultaneous masking” (USSM, Egan and Hake, 1950; Oxenham and Plack, 1998; Bacon *et al.*, 1999; Yasin and Plack, 2005). The USSM refers to the nonlinear (excessive) growth of masking for maskers lower in frequency than the target, compared to maskers close to the CF of the target. Oxenham and Moore (1995) proposed an analogy between USSM and simultaneous masking additivity. Particularly, they suggested that a 3-dB increase in masker level in an USSM experiment may have the same effect at the CF of the target as combining two equally effective and lower-frequency maskers in an additivity experiment. Thus, in analogy with the additivity-of-simultaneous-masking situation, linear masker processing should result in USSM while compressive masker processing should result in linear growth of masking. Noteworthy, this analogy between USSM and simultaneous masking additivity relies on the assumption that nonlinear interactions between maskers (such as suppression) are ruled out in the additivity experiment.

With respect to the MOCR, there is at least one reason why simultaneous masking additivity for the Gaussian stimuli used in the present study may differ from that observed in the cited studies of masking additivity (Humes *et al.*, 1992) and USSM (Oxenham and Plack, 1998; Bacon *et al.*, 1999; Yasin and Plack, 2005). All those studies involved relatively long maskers and presented the target at the temporal center of the masker, while the Gaussians used in the present study had a very short duration (<10 ms). The delay between masker and target onsets in the past studies was at least about 100 ms,² which is longer than the onset delay of the MOCR. Therefore, compression of the target and thus excess masking (as well as the growth rate of USSM) can be expected to be lower in those studies than in the present study using short stimuli. This expectation is based on the assumption that the MOCR effect elicited by the spectrally remote and long maskers reached the target. Accordingly, Lilaonitkul and Guinan (2009) reported a relatively broad spectral spread (from -1.5 octaves below to $+0.5$ octaves above the CF for an ipsilateral elicitor) of the MOCR effect around the frequency of the elicitor. By using short Gaussian stimuli, we aimed at measuring simultaneous masking additivity while excluding any effect of the MOCR.

We tested combinations of maskers with frequencies below the target (lower-frequency maskers) and with frequencies both below and above the target (spectrally flanking maskers). The main questions to be addressed were: Can the amount of excess masking for the different combinations of flanking and lower-frequency maskers be explained by a model assuming linear processing of the maskers and compression of the target? If not, what modifications of the assumptions on cochlear processing of off-CF stimuli are required to explain the results? Are the results consistent with the idea that a short stimulus avoids activation of the MOCR, resulting in more compression and, thus, more excess masking compared to long stimuli? Given that excess masking can be influenced by several factors including suppression, off-frequency listening and availability of detection cues, we carefully attempted to estimate their potential influences.

The comparison between flanking and lower-frequency maskers is interesting in the light of the results from recent psychoacoustical nonsimultaneous masking experiments that indicate partial compression of lower-frequency maskers spectrally separated from the target by almost an octave (Lopez-Poveda and Alves-Pinto, 2008) or even lower (Plack and Arifianto, 2010). Note that for such spectral configurations, most physiological studies on BM vibration from the literature suggest linear processing of the maskers at the target CF. The cited authors suggested that the partial compression of the maskers inferred from their psychoacoustical data might not result from the CF-specific outer hair cell activity but rather proposed the compressive CF-independent nonlinearity of the inner hair cells (Cheatham and Dallos, 2001; Lopez-Poveda and Eustaquio-Martin, 2006) as a candidate. With respect to simultaneous masking additivity, compression of the spectrally remote maskers may result in reduced excess masking relative to the prediction based on linear processing of the maskers. Moreover, if the inner hair cells are actually responsible for this compression, then this

should affect all maskers similarly, irrespective of their frequency distance from the target. In other words, excess masking would be expected to be similarly reduced when combining either lower-frequency maskers or lower- and higher-frequency maskers.

II. METHODS

A. Listeners

Five normal-hearing (NH) listeners participated in the experiments. All listeners had thresholds of 15 dB hearing level (HL) or lower at octave frequencies from 125 to 8000 Hz (ANSI, 1996) and had previous experience in psychoacoustical tasks. All listeners except NH14 participated in the preceding study on nonsimultaneous masking additivity (Laback *et al.*, 2011).

B. Stimuli and apparatus

All masker and target stimuli were Gaussian-shaped tones defined by

$$s(t) = \sqrt{\Gamma} \cdot \sin\left(2\pi f_0 t + \frac{\pi}{4}\right) \cdot e^{-\pi(\Gamma t)^2}, \quad (1)$$

where f_0 is the tone frequency and $\Gamma = \alpha \cdot f_0$. For a given f_0 , the shape factor of the Gaussian window, α , controls the duration and bandwidth of $s(t)$. In Laback *et al.* (2011), a single value of f_0 (4 kHz) was used and α was fixed to 0.15, which provided the Gaussians with a constant equivalent rectangular bandwidth ($\text{ERB}_{\text{Gauss}} = \Gamma = 600$ Hz) and a constant equivalent rectangular duration ($\text{ERD}_{\text{Gauss}} = \Gamma^{-1} = 1.7$ ms). In the present study, f_0 varied depending on the spectral separations between maskers and target. By keeping α constant, Γ would have varied with f_0 , which would have caused the spectro-temporal shape of the stimulus to vary. In order to avoid such a variation, we decided instead to fix the Γ value to 600 so as to keep $\text{ERB}_{\text{Gauss}}$ and $\text{ERD}_{\text{Gauss}}$ as in Laback *et al.* By introducing the $\pi/4$ phase shift, the energy of the signal was independent of f_0 .

The total signal duration, as given by the numerical support, was 9.6 ms. The sound pressure level (SPL) of a Gaussian was specified by measuring the sound pressure level (SPL) of a long-lasting sinusoid having the same frequency and amplitude as the carrier tone of the Gaussian.

A personal computer system was used to control the experiments and generate the stimuli. Stimuli were output at a sampling rate of 48 kHz and a 24-bit resolution with an external D-A converter (AD/DA 2402, Digital Audio Denmark), passed through an attenuator (PA4, Tucker-Davis Technologies, TDT) and a headphone amplifier (HB6, TDT), and routed to the left-ear side of a circumaural headphone (HDA200, Sennheiser). The experiments were performed in a double-walled, sound-attenuated booth.

C. Procedure

Detection thresholds were measured using an adaptive three-interval forced-choice task. The target was presented randomly in one of the three intervals. In absolute threshold

measurements, the other two intervals were silent. In masked threshold measurements, all three intervals contained the masker(s). The listeners had to indicate which interval sounded different from the other two by pressing one of three buttons of a keyboard. Each 200-ms interval was visually indicated on a computer screen with a between-interval gap of 200 ms. The stimuli were presented in the temporal center of the interval. Response feedback was provided after each trial by visually highlighting the interval containing the target. In the adaptive procedure, the target level was varied using the three-down one-up rule, estimating the 79.4% point on the psychometric function (Levitt, 1971). The initial step size was 5 dB and was halved after the second reversal. A run was terminated after 12 reversals and the threshold was calculated by averaging the target level over the last eight reversals.

The different experimental conditions were presented in a quasi-randomized order for each listener, as described in Laback *et al.* (2011). The final threshold for each condition and listener was determined by computing the arithmetic mean of, on average, 5.5 valid runs. The method for selecting valid runs and tracking for learning effects was the same as described in Laback *et al.* (2011). The total testing time for one listener was seven to eight hours. All other aspects of the procedure were the same as in Laback *et al.* (2011).

D. Experimental conditions

The target and the maskers were presented simultaneously. The target had a frequency of 5611 Hz. The four maskers were separated in frequency from the target by the following frequency values defined in the ERB scale (Glasberg and Moore, 1990): -7 ERBs (M1), -5 ERBs (M2), -3 ERBs (M3), and $+3$ ERBs (M4). The corresponding frequency values in Hz are provided in Table I. Three maskers had lower frequencies and one masker had a higher frequency relative to the target (see inset at the top of Fig. 1). Based on pilot experiments, the frequency separations were chosen so as to keep minimum overlap between the maskers and to avoid exceeding a comfortable sound level for the most distant maskers. An asymmetric configuration (three lower-frequency maskers vs one higher-frequency masker) was required because of the narrower spread of

spectral masking towards lower frequencies than towards higher frequencies (e.g., Egan and Hake, 1950). All masker-to-target frequency separations were chosen so that the maskers should be processed linearly at the CF of the target according to results from the majority of physiological BM measurements (e.g., Ruggero, 1992; Ruggero *et al.*, 1997).³

Simultaneous masking may involve the perception of combination tones. The frequency separations were chosen, however, to minimize the potential influence of combination tones. To test their potential contribution, we explored in a pilot test the potentially most critical condition where the target is closest and above the masker in frequency (Greenwood, 1971), i.e., masker M3. The masked threshold for M3 was tested with and without a continuous bandpass-filtered white noise. The noise had low and high cutoff frequencies of 50 and 2530 Hz⁴ and a level of 40 dB SPL, sufficient to mask the cubic difference tone (Greenwood, 1971). This pilot test was performed with listeners NH2 and NH23. Repeated measurements of the masked threshold for the conditions with and without noise in a balanced order showed no systematic difference in the masked thresholds. To avoid potentially confounding effects of the background noise, we decided not to use the background noise in the main experiments.

In the first stage of the experiment, the absolute threshold of the target was measured. In the second stage, the level of each masker was assessed which was necessary to produce approximately 10 dB of masking of the target. This was achieved using an iterative approach where the masker level was adjusted after each run in order to reach the desired masked threshold, as used in Laback *et al.* (2011).⁵ After the final masker level had been determined, further measurements at this level were performed in a separate test session. These measurements represent the reported masked thresholds for the single-masker conditions. In the third stage, the main experiment, masked thresholds were measured for selected combinations of maskers. These combinations were M2M3, M3M4, M1M2M3, M2M3M4, and M1M2M3M4. After completion of the main experiment, the single-masker conditions were retested to check for learning effects. No learning was observed. The data presented below therefore correspond to the mean overall data collected. For more details about the methodology, see Laback *et al.* (2011).

III. RESULTS AND DISCUSSION

A. Single maskers

Table I presents individual absolute thresholds of the target (second column) and the masker levels necessary to produce approximately equal amounts of masking of the target (i.e., masker levels for which the target level at threshold was about 10 dB above absolute threshold; columns 3–6). The level required is higher for masker M4 ($+3$ ERBs) than for M3 (-3 ERBs), M2 (-5 ERBs), and M1 (-7 ERBs). This asymmetry in equally effective masker levels is consistent with the well-known finding that the spread of spectral masking extends further towards higher frequencies than towards lower frequencies (e.g., Egan and Hake, 1950; Moore, 2003).

Figure 1 presents the individual and mean masking data in separate panels. The results are reported as amounts of

TABLE I. Absolute thresholds for the 5611-Hz target and masker levels necessary to produce about 10 dB of masking of the target, for each of the four maskers.

Subject	Absolute threshold of target (dB SPL)	Equally effective masker level (dB SPL)			
		M1 (-7 ERBs) (2521 Hz)	M2 (-5 ERBs) (3181 Hz)	M3 (-3 ERBs) (4000 Hz)	M4 ($+3$ ERBs) (7836 Hz)
NH2	40.2	71.9	67.6	65.3	77.2
NH19	34.2	61.9	59.6	56.3	66.2
NH23	32.2	65.9	62.6	61.3	71.2
NH25	33.2	65.9	62.6	61.3	71.2
NH14	37.2	67.9	65.6	63.3	74.2
Mean	35.4	66.7	63.6	61.5	72
SD	3.3	3.6	3.1	3.3	4.1

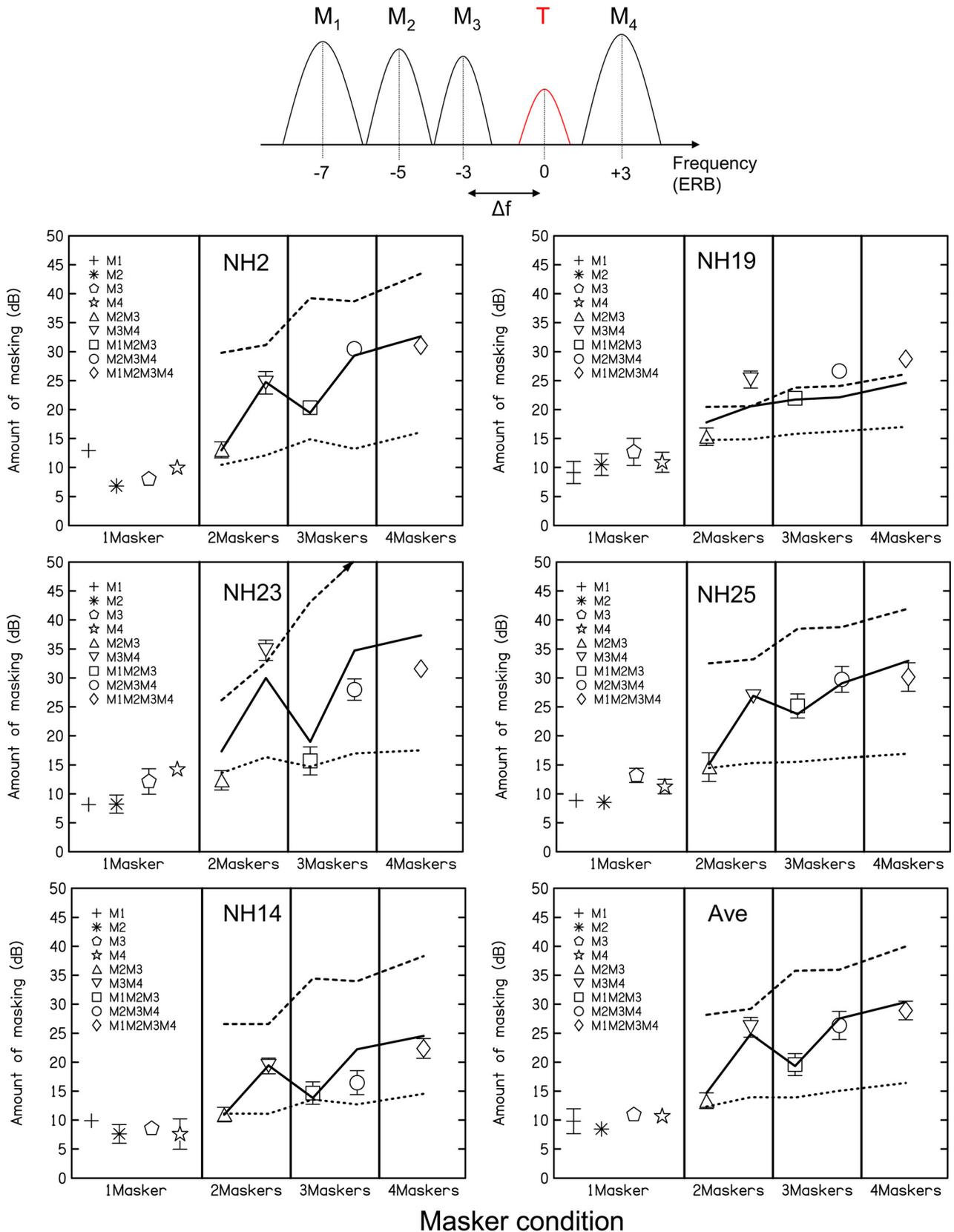


FIG. 1. (Color online) Individual and mean amounts of masking obtained for each masker condition (see symbols in the legend). Each panel shows the results for one listener, except for the bottom-right panel showing the mean results. Error bars indicate 95% confidence intervals. In many cases the error bars are smaller than the symbols. The lines show the amounts of masking predicted by three models of masking additivity: (1) linear additivity (i.e., no excess masking; dotted line), (2) “standard model of simultaneous masking additivity” that assumes compression of the target and linear processing of the maskers (dashed line; for listener NH23 the prediction for the quadruple exceeds the border of the plot and is indicated by an arrow), and (3) a model assuming weighted summation of the lower-frequency maskers (solid line). The weighting factors for each of the maskers (k_n) were optimized to best fit the data (see text for details). The best-fitting k_n values for each listener and the mean data are shown in Table II. The inset at the top of the figure schematically illustrates the spectral arrangement of the stimuli.

masking, defined as the difference between the masked and absolute thresholds of the target. The amounts of masking for the single maskers are shown in the leftmost column of each panel. On average, they amounted to 9.9 dB (M1), 8.4 dB (M2), 10.9 dB (M3), and 10.7 dB (M4). The difference between absolute and masked thresholds was statistically significant for all maskers [repeated-measures analysis of variance, RM ANOVA: $F_{3,83} = 72.8$, $p < 0.001$; Tukey's HSD *post hoc* test: $p < 0.001$].

B. Multiple maskers

The amounts of masking for the combinations of two, three and four maskers are shown in Fig. 1 (see legend for meaning of different symbols). The amounts of masking were quite similar across listeners, as indicated by the small 95% confidence intervals in the bottom-right panel. Thus, we focus below on the description of the mean data. An RM ANOVA was performed with the factor masking condition, comparing all single and multiple masker configurations. The main effect of masking condition was highly significant ($F_{8,179} = 149.5$, $p < 0.001$). The significance of the differences between the factor levels was analyzed with a Tukey's HSD *post hoc* test.

For the pair of lower-frequency maskers (M2M3), the mean amount of masking was about 3 dB higher than that obtained for each masker alone (M2M3 vs M2: $p < 0.001$; M2M3 vs M3: $p = 0.24$). This is in line with the linear additivity prediction (see dotted line), that is, almost no excess masking (0.9 dB). In contrast, for the pair of flanking maskers (i.e., combination of lower- and higher-frequency maskers, M3M4), the difference in masking to the mean value for the corresponding single maskers amounted to the much larger value of 16.0 dB ($p < 0.001$ for both comparisons) and excess masking was 12.2 dB. The difference in masking between conditions M2M3 and M3M4 was 12.7 dB ($p < 0.001$).

Adding a third lower-frequency masker (M1) to the pair of lower-frequency maskers (M2M3) significantly increased the amount of masking (5.7 dB; $p < 0.001$). In contrast, adding a lower-frequency masker (M2) to the pair of flanking maskers (M3M4) had no significant effect on the amount of masking (0.33 dB; $p = 1$). Excess masking was much greater for the triplet of flanking maskers (M2M3M4, 11.2 dB) than for the triplet of lower-frequency maskers (M1M2M3, 4.7 dB).

Adding a fourth masker affected the amount of masking only when a higher-frequency masker was added to a lower-frequency triplet (M1M2M3M4 – M1M2M3, 9.9 dB; $p < 0.001$; M1M2M3M4 – M2M3M4, 2.6 dB; $p = 0.13$). Excess masking for the quadruple was 12.5 dB.

Overall, these comparisons suggest that combining flanking maskers produces strong excess masking while combining (two or three) lower-frequency maskers produces only very little excess masking.

C. Comparison to the literature

We confined the comparison to the literature to (1) studies involving sufficient frequency separation between maskers and target so as to assume linear processing of the maskers at the CF of the target and (2) sensation levels of

the target comparable to our study. Humes *et al.* (1992) tested a configuration similar to our condition M3M4 using bandpass-filtered noise maskers spectrally flanking a 2000-Hz target. They tested several masker-to-target spectral separations. On average, for small spectral separations (i.e., < 200 Hz) leading to spectral overlap with the target, excess masking was close to zero. For larger spectral separations, excess masking amounted to 4.4 dB. This amount of excess masking is considerably less than that observed in condition M3M4 of the present study. The smaller amount of excess masking in Humes *et al.* (1992) is consistent with the idea that the delay between masker and target onsets in their study (100 ms) caused the MOCR to reduce BM compression of the target. Note that for the spectral separations between maskers and target in that study (less than one octave for the lower-frequency maskers), the masker-induced MOCR would still have affected the target (Lilaonitkul and Guinan, 2009).

In a second experiment, Humes *et al.* (1992) tested a condition similar to our condition M2M3, namely a pair of lower-frequency maskers. One masker was a 105-Hz pure tone and the other masker was a bandpass-noise (275–1075 Hz). These stimuli were designed to minimize confounding effects of combination tones and envelope cues, as might have been involved in earlier studies testing pairs of lower-frequency maskers (Zwicker and Herla, 1975; Lutfi, 1983). On average across the various target frequencies tested, excess masking was 3.9 dB, which is slightly more than for our condition M2M3.

Excess masking for lower-frequency maskers is thought to be related to USSM (see Introduction). For target-masker frequency ratios and target levels approximately comparable to our condition M2M3, masking growth rates of about 2.0 (Oxenham and Plack, 1998; Bacon *et al.*, 1999) to 2.5 dB/dB (Yasin and Plack, 2005) have been reported. Considering the average rate of 2.2 dB/dB across these studies and assuming intensity summation of the individual masker excitations would lead to excess masking for a masker pair of $(3 \times 2.2) - 3 = 3.6$ dB.⁶ This is slightly more than the 0.9 dB excess masking obtained for condition M2M3. For a masker triple the same growth rate would lead to excess masking of $(4.8 \times 2.2) - 4.8 = 5.8$ dB, which is close to the 4.7 dB excess masking observed for condition M1M2M3.

In summary, excess masking for lower-frequency Gaussian maskers appears to be in the same order as that reported in the literature on simultaneous masking additivity and USSM. However, excess masking for spectrally flanking Gaussian maskers appears to be larger than in a previous study applying a time delay between the masker and target onsets that may have involved post-cochlear adaptation effects such as the MOCR.

D. Factors potentially influencing excess masking

Assuming linear processing of the maskers and compression of the target at the target CF, the standard model of simultaneous masking additivity predicts that excess masking is similar for a pair of lower-frequency maskers and a pair of flanking maskers. However, our results showed much stronger excess masking for the flanking maskers than for the lower-frequency maskers.

There are a number of factors that could contribute either to increased or decreased excess masking relative to the prediction of the standard model.

1. Factors resulting in increased excess masking

As noted by Humes *et al.* (1992), excess masking for flanking maskers can potentially be increased due to off-frequency listening. Namely, in the single-masker conditions the listeners can detect the target in a filter spectrally remote from the masker, providing improved target detectability. This off-frequency listening provides no advantage when the two maskers representing the flanking-masker condition are combined. Although off-frequency listening effects are generally relatively small at low target levels, we estimated their contributions to the amount of excess masking for our condition with flanking maskers (M3M4, see Appendix). The outcome of this analysis is that off-frequency listening effects may have led to overestimation of the amount of excess masking for condition M3M4. However, the amount of overestimation was predicted to be no more than 0.5 dB. Thus, the “corrected” excess masking amounts to $12.2 - 0.5 = 11.7$ dB,⁷ which is still considerably more than that obtained for condition M2M3 (for which off-frequency listening provides no advantage).

Any type of dip-listening cue available in the single-masker conditions but not in the multiple-masker condition can potentially result in overestimation of excess masking (Moore, 1985). For the stimuli of the present study, however, this explanation is unlikely because target and maskers had the same envelopes and were gated simultaneously.

The perception of combination tones can also lead to overestimation of excess masking by providing a detection cue in the single-masker conditions which could be masked in the combined-masker condition (Moore, 1985). However, for the spectral separations chosen in the present study, none of the most prominent combination tones (the cubic and the simple difference tones, see Greenwood, 1971) was critical with respect to such an effect. Moreover, pretests on the most critical condition (M3) showed no effect of adding a background noise used to mask combination tones (see Sec. II D).

Finally, the availability of spectral shape cues could have played a role. For the single- and multiple-masker conditions with maskers on the lower-frequency side only (e.g., M2M3) where the target is the highest spectral component, the listeners could have had access to salient spectral shape cues, e.g., changes in the spectral tilt or the spectral center of gravity. However, such cues are less likely to have been accessible in the conditions with flanking maskers (e.g., M3M4). This could have enhanced the amount of excess masking in condition M3M4.

2. Factors resulting in decreased excess masking

Between-masker suppression effects could decrease excess masking relative to the prediction of the standard model if one masker reduces the “internal” response to another masker (Humes and Jesteadt, 1989). This could occur particularly for the lower-frequency maskers and thus

explain the lack of excess masking for condition M2M3. In order to estimate the potential contribution of suppression, we used a nonlinear model of the auditory periphery as described in Plack *et al.* (2002). This model features the dual resonance nonlinear filter introduced in Meddis *et al.* (2001) and has been shown to be able to predict several masking and suppression data (Plack *et al.*, 2002; Recio-Spinoso and Lopez-Poveda, 2010). First, we simulated the suppression of masker M3 by M2, which amounted to 2.7 dB.⁸ Second, we predicted the masked threshold of the target in the presence of M2 and M3 either with their original equally effective levels (63.6 and 61.5 dB SPL, respectively) or with M3 reduced by 2.7 dB. The predicted masked threshold of the target was only 1.3 dB lower with suppressed M3 than with original masker levels. This indicates that suppression of M3 by M2 had only a minor effect on effective masking and thus does not appear to be a plausible explanation for the small amount of excess masking for combination M2M3. Additionally, note that the estimated small amount of suppression for equal levels of suppressor (M2) and suppressee (M3) is consistent with several earlier results (e.g., Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990).

Partial spectral overlap between adjacent maskers could lead to uncontrolled interference effects and thus reduce the amount of excess masking. In order to study the potential effect of the spectral overlap for condition M2M3, we tested two additional masker pairs avoiding spectral overlap on five listeners (NH19 was replaced by a new listener). For both pairs the closer masker was identical to M3 (−3 ERBs *re* target), while the more remote masker was set to either −7 or −9 ERBs (*re* target). As a control condition, M2M3 was also included. All other aspects were identical to the main experiment. The mean amount of excess masking was 1.0, 0.0, and 2.3 dB, respectively, for the M2M3, −7 ERBs and −9 ERBs conditions. Thus, the spectral overlap of the maskers was not responsible for the lack of excess masking in condition M2M3.

Decreased excess masking could be due to effects of the maskers on the shape of the I/O function for the target. Both physiological (Ruggero *et al.*, 1992; Rhode and Recio, 2001) and psychophysical studies (e.g., Yasin and Plack, 2007) showed that the presence of a suppressor can linearize the I/O function, i.e., reduce the compression, for a target at CF. Considering the maskers of the present study as suppressors, the data could be explained by assuming a stronger linearization effect for lower-frequency suppressors than for higher-frequency suppressors. While available data appear to show a trend towards such a difference (e.g., Rhode and Recio, 2001; Yasin and Plack, 2007), the effects depend in a complex manner on the specific frequency distance between the suppressor and the target CF and the specific levels of target and suppressor. Moreover, it is not clear if, or under which conditions, the linearization effects of two (or more) maskers add up. The available literature does not provide sufficient information to allow estimating the linearization effects for the stimulus configurations of the present study. We therefore used the nonlinear cochlear filter model described in Plack *et al.* (2002) to quantitatively predict the target I/O functions in presence of the masker combination M2M3.

The results showed only marginal linearization in condition M2M3, suggesting that linearization was not the reason for reduced excess masking for that condition.

Thus far, we assumed that all maskers in the present study are processed linearly at the CF of the target. It is possible, though, that the lower-frequency maskers are at least partially compressed at the CF of the target. Accordingly, there are some recent indications of partial compression of maskers almost an octave below the target (Lopez-Poveda and Alves-Pinto, 2008) or even lower (Plack and Arifianto, 2010). Consequently, the summed response to the combined maskers at the target CF may be smaller than in case of linear processing of the maskers. This would in turn lower the required increase in target level for the combined *re* single masker conditions in order to maintain the same target-to-masker ratio after cochlear compression, leading to reduced excess masking. In the next section, we quantitatively predict excess masking for our experimental conditions by considering both intensity summation (standard model) and weighted intensity summation of excitations from spectrally remote maskers.

E. Modeling

First, the results for the combined maskers were predicted by the standard model of simultaneous masking additivity. Because this model assumes that the excitations of the individual maskers at the target CF are summed in intensity units, the same framework as that used to predict the additivity of forward masking (Plack *et al.*, 2006; Plack *et al.*, 2008; Laback *et al.*, 2011) can be applied, with the conceptual difference that the masker effects are summed simultaneously, i.e., jointly, rather than by means of a temporal integration window. A measure of the masking effect, E , can be taken as the target intensity at masked threshold after cochlear compression,

$$E = f(S), \quad (2)$$

where S is the target intensity at threshold (in dB). Following Plack *et al.* (2006), the cochlear I/O function was modeled as a third-order polynomial defined by

$$f(x) = ax^3 + bx^2 + cx, \quad (3)$$

where x is the input signal intensity (in dB SPL) and a , b , and c are coefficients (note that the intercept of the function is not constrained by the data and does not affect the predictions of the model). Assuming that the effects of the maskers (M_1, \dots, M_n) add linearly,

$$E_{COMB} = E_{M1} + \dots + E_{Mn}, \quad (4)$$

where $E_{M1} \dots E_{Mn}$ are the masking effects of each masker and E_{COMB} is the combined masking effect. Substituting from Eq. (2) and solving for S , including the conversion from dB to intensity units, gives

$$S_{COMB} = f^{-1}(10 \log(10^{f(S_{M1})/10} + \dots + 10^{f(S_{Mn})/10})), \quad (5)$$

where $S_{M1} \dots S_{Mn}$ are the target intensities at threshold (in dB) in the presence of the individual maskers and S_{COMB} is the target intensity at threshold (in dB) in the presence of n maskers combined. Using this equation in combination with Eq. (3) as the function f , the thresholds for the single maskers ($S_{M1} \dots S_{Mn}$) were used as the inputs of the model and the thresholds in the presence of the combined maskers (S_{COMB}) were predicted. As estimates of the function f , we used the I/O functions derived in Laback *et al.* (2011) for nonsimultaneous Gaussian maskers assuming that (1) BM compression of the target is similar in nonsimultaneous and simultaneous masking, as suggested in Yasin and Plack (2005), and (2) the shapes of the I/O functions are approximately comparable at the two frequencies 5611 and 4000 Hz. For listener NH14, who did not participate in the study of Laback *et al.* (2011), we used the mean I/O function derived from the other four listeners.

The dashed lines in the panels of Fig. 1 show the model predictions. The model gives a reasonable account of masking additivity only for condition M3M4 (on average, excess masking in that condition is overestimated by about 3 dB). For all other conditions, the model systematically overestimates masking additivity by at least 10 dB. The only exception is listener NH19, for whom the prediction error is much smaller. This listener showed the steepest—thus the least compressive—I/O function in Laback *et al.* (2011). The reasonable prediction for condition M3M4 is consistent with the assumption that the strong BM compression observed in nonsimultaneous masking also applies to simultaneous masking, using short Gaussian stimuli in both cases.

The predictions also indicate that some mechanism that is not accounted for in the model limits excess masking for the lower-frequency maskers. As described above, it is conceivable that the combined response to the maskers at the target CF is less than that corresponding to intensity summation, which would result in less excess masking than predicted by the standard model. Because the maskers themselves are not included in the model, their processing cannot directly be incorporated into the model. However, the consequence of nonlinear masker processing on the summation of individual masker effects ($E_{M1} \dots E_{Mn}$) can be incorporated into the model by weighting their measures ($f(S_{M1}) \dots f(S_{Mn})$)

$$S_{COMB} = f^{-1}(10 \log(10^{[f(S_{M1})+a(k_1-1)]/10} + \dots + 10^{[f(S_{Mn})+a(k_n-1)]/10})). \quad (6)$$

The factors $k_1 \dots k_n$ represent the weightings⁹ for the individual maskers ($M_1 \dots M_n$) in the summation process. In order to obtain intensity summation of masking effects when all k values are 1 and to obtain no summation at all (i.e., no excitation difference between single and multiple maskers) when all k values are zero, the model parameter a has to be set depending on the number of maskers: it has the value 3.0 for two maskers, 4.77 for three maskers, and 6.0 for four maskers.¹⁰ By this dependency of the exponents of the individual masker terms on the total number of maskers (and on k), the model mimics the variable contribution of the individual maskers in the summation process. In other words, it controls the efficiency of the summation of masker-induced

excitations at the target CF. Note that this formulation predicts maximum excess masking in case of intensity summation ($k=1$) and linear masking additivity (no excess masking) if k equals the slope of the I/O function for the target ($f(x)$). Note the difference of this model to the nonsimultaneous masking model. For the latter, the individual maskers are processed *before* summing their effects and thus excess masking depends on the I/O function of the target only, not on the particular processing of the maskers (see Appendix of [Oxenham and Moore, 1995](#)).

We determined the values of k_1 , k_2 , k_3 , and k_4 that best predict the overall pattern of the data in a two-stage approach. In the first stage, k_2 , k_3 , and k_4 were optimized to best predict the results for the conditions M2M3 and M3M4 by minimizing the sum of the squared deviations between predictions and data (in dB). In the second stage, k_2 , k_3 , and k_4 were fixed to the optimum values found in the first stage and k_1 was optimized to best predict the remaining conditions M1M2M3, M2M3M4, and M1M2M3M4. The reason for using this two-stage approach was that the conditions with more than two maskers could potentially involve more complex interaction effects such as suppression as compared to the masker pairs. Thus, we attempted to minimize the influence of such effects on the predictions for the masker pairs M2M3 and M3M4. The optimum k -values determined for the individual listeners and for the mean data are listed in Table II. The model predictions for the individually “optimized” k -values are shown in Fig. 1 with solid lines.¹¹ The results show that the data can be best predicted assuming low to moderate weighting of the lower-frequency maskers (M1, M2, M3) and high weighting of M4 in the summation process. The optimized k -values of some listeners vary nonmonotonically across the three lower-frequency maskers with larger values for M2 than for M1 and M3. One possible interpretation of the k -values is to consider them as reflecting the amount of compression the individual maskers are subjected to at the target CF. Based on this interpretation, the nonmonotonic variation of k -values appears implausible from a physiological point of view. Even though a large amount of within and across-listener variability appears to be typical for psychophysical estimates of cochlear compression (e.g., [Rosengard et al., 2005](#)), this unexpected behavior may suggest that some other factors besides compression influenced the summation of masker excitations. Finally, it should be noted that our approach to freely vary the individual k -values for all four maskers and, thus, to somehow over-fit the model, was motivated by the goal of identifying

a configuration of weighting factors in the summation process that best explained the data for all masker combinations rather than obtaining the best predictions of masked thresholds.

IV. SUMMARY AND CONCLUSIONS

In this study we reported measurements of the additivity of simultaneous masking for several combinations of short Gaussian-shaped tones. According to the majority of the published physiological BM measurement studies we are aware of, all maskers were sufficiently spectrally separated from the target so as to be considered as being processed linearly at the CF of the target. For such a configuration, the “standard” model of simultaneous masking additivity ([Oxenham and Moore, 1995](#)) predicts excess masking. This model considers that the target is compressed and assumes that the 3-dB increase of masker-induced excitation resulting from the addition of a second masker requires more than a 3-dB increase of the target level relative to the single-masker condition to maintain the threshold criterion. The results of the present study for a pair of flanking maskers (condition M3M4) indeed showed strong excess masking (mean value: 12.2 dB), in agreement with the predictions of the standard model using I/O functions derived from an experiment on nonsimultaneous masking additivity with the same type of stimuli and listeners ([Laback et al., 2011](#)). In that study, excess masking was found to be stronger than in previous nonsimultaneous masking studies using comparable target levels but longer masker durations. The stronger excess masking was attributed to the much shorter interval between the onset of the first masker and the onset of the target in [Laback et al. \(2011\)](#) than in previous studies. It was argued that if the between-stimuli interval exceeds the onset delay of the MOCR, the MOCR reduces the cochlear gain (and thus probably also the amount of compression) for the target, resulting in less excess masking. Similarly, the amount of excess masking for the flanking *simultaneous* maskers of the present study was found to be much stronger than in a previous study ([Humes et al., 1992](#)) that used a comparable spectral configuration but involved a delay between the masker and target onsets. Because this delay exceeded the MOCR onset delay, the activation of the MOCR seemed to have reduced the amount of excess masking in [Humes et al. \(1992\)](#) as compared to the present study. Together, the results suggest similar amounts of excess masking in simultaneous masking with spectrally flanking maskers and in nonsimultaneous masking if the activation of the MOCR is avoided. It should be noted that the effects attributed here to the MOCR might also be due to some other post-cochlear adaptation mechanism.

We have discussed the role of factors potentially causing overestimation of excess masking for flanking maskers. We provided arguments for why it is unlikely that off-frequency listening, envelope cues, or combination tone perception played an important role. Although the restricted access to spectral-shape cues for the flanking-masker condition could have contributed to excess masking, an important influence of such spectral cues does not appear likely, given

TABLE II. Weighting factors k_n corresponding to each of the maskers M_n that result in model predictions best fitting the experimental data.

Listener	k_1	k_2	k_3	k_4
NH2	0.3	0.3	0.2	1
NH19	1	0.1	1	1
NH23	0.1	0.1	0.8	1
NH25	0.6	0.1	0.2	0.9
NH14	0.2	0.1	0.2	0.9
Mean	0.44	0.14	0.48	0.96
SD	0.36	0.09	0.39	0.05

that the amount of excess masking was well-predicted by the “standard” model of masking additivity.

In contrast to the flanking maskers, the results for a pair of lower-frequency maskers (condition M2M3) showed essentially no excess masking (mean value: 0.9 dB). Similarly, for combinations of more than two lower-frequency maskers, very little excess masking was observed. These results are inconsistent with the standard model of simultaneous masking additivity that does not predict any difference between lower-frequency and flanking maskers. We discussed a number of factors potentially causing reduced excess masking for the lower-frequency maskers. We showed that between-masker suppression and between-masker spectral overlap appear to be unlikely explanations. We considered also the linearization of the target I/O function by the suppressive effect of the maskers. However, the available literature data appear to be too limited to allow an interpretation of our data, particularly with respect to the dependency of the linearization effect on the frequency and level relationships between suppressor and target. Therefore, we performed a model simulation of the linearization effect using a nonlinear filter model, showing only marginal linearization of the target I/O function and, thus, providing no support for the linearization explanation.

We then questioned the assumption of linear processing of the maskers, inspired by some recent psychophysical studies suggesting partial compression of lower-frequency maskers (Lopez-Poveda and Alves-Pinto, 2008; Plack and Arifianto, 2010). Because partial compression of the lower-frequency maskers could explain reduced excess masking, we reran a modification of the standard model that incorporated a weighting of the contributions of the individual maskers in the summation process as free parameters. The data were best predicted when assuming high weighting of the higher-frequency masker and low to moderate weighting of the lower-frequency maskers. One interpretation of the weights in the summation process is that they reflect the amount of compression the individual maskers are subjected to at the target CF. Plack and Arifianto (2010) suggested that the compression of lower-frequency maskers may be due to the compressive nonlinearity of the inner hair cells (Cheatham and Dallos, 2001; Lopez-Poveda and Eustaquio-Martin, 2006). Because such a mechanism appears to affect all CFs similarly, it may explain to some extent the lack of excess masking for the lower-frequency maskers but it can hardly explain the strong excess masking observed for flanking maskers. Thus, currently we cannot offer a physiologically plausible explanation for the data. It should be considered, though, that physiological measurements of on- and off-frequency I/O functions without and with suppressors are mostly based on stimuli which were much longer than the very short stimuli of the present study. The results might differ for such short stimuli. To that end, collecting physiological measurement of I/O functions for short stimuli placed below, at, and above CF may provide insight.

Finally, our results on the excess masking for the lower-frequency maskers were found to be roughly consistent with data on the USSM for similar target-masker frequency ratios and target levels (Oxenham and Plack, 1998; Bacon *et al.*, 1999; Yasin and Plack, 2005). Both excess masking and

USSM are smaller than expected from linear processing of the maskers and compression of the target. We assume that increasing the level of a lower-frequency masker (USSM) by 3 dB and combining two lower-frequency maskers (masking additivity) both cause the same excitation change at the CF of the target. Thus, we can also assume that the mechanism(s) limiting excess masking for lower-frequency maskers are the same as those that limit the growth rate of USSM. In psychophysical (Yasin and Plack, 2003, 2007) and physiological (Ruggero, 1992) studies, it has been suggested that for mid-level targets, lower-frequency maskers may reduce the USSM by linearizing the I/O function. To our knowledge, an explanation in terms of weighted summation such as resulting from compression of lower-frequency maskers has not yet been considered in the literature.

ACKNOWLEDGMENTS

We thank Gwenael Toupin for programming the software routines required for the experiments, Katharina Egger for collecting and preparing some of the data, and Christian Kaseß for proofreading of an earlier version. We also thank Dr. Larry Humes for stimulating ideas. Parts of the work were carried out while the first author was a fellow at the Hanse Institute for Advanced Study, Delmenhorst, Germany. The work was partly supported by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF, project MA07-025), the Wissenschaftlich-Technologische Zusammenarbeit (WTZ, project Amadee), the Agence Nationale de la Recherche (ANR, project “SenSons”), and the Austrian Science Fund (FWF, START project Y 551-N13).

APPENDIX

The measurement of masked thresholds for the single-masker conditions may have involved off-frequency listening effects: the listeners could have used auditory filters remote from the target frequency to detect the target. This could provide a detection advantage in the single-masker conditions but not in the flanking masker condition (M3M4), which would in turn affect the estimated amount of excess masking. In order to quantitatively estimate the effect of off-frequency listening for maskers M3 and M4, we simulated the signal-to-masker level ratios at the output of the nonlinear model of the auditory periphery as described in Plack *et al.* (2002) at center frequencies surrounding the target. Twenty-two filters equally spaced on the ERB-rate scale within the range 5015 to 6340 Hz and four additional filters for condition M4 were calculated. The signal-to-masker ratios were determined for masker and target levels as used in the experiment, taking the mean data across listeners. The absolute threshold was taken into account as a noise floor. As an estimate of the absolute threshold, we used the shape of the minimum audible pressure (MAP) correction curves (ISO recommendation R. 389), up-shifted in level by the difference in mean absolute threshold for a Gaussian at 5611 Hz and the MAP curve at that frequency. Figure 2 shows the signal-to-masker ratios (in dB) at the outputs of the auditory filters as a function of the filters’ center

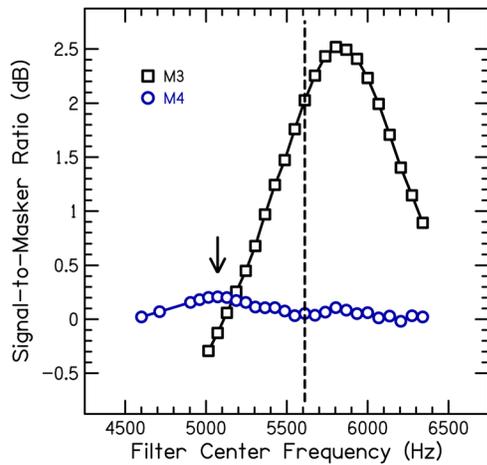


FIG. 2. (Color online) Estimated signal-to-masker level ratios at the outputs of simulated auditory filters as a function of the filters' center frequency for maskers M3 (squares) and M4 (circles). The vertical dashed line indicates the target frequency (5611 Hz). The curve for M3 has the maximum at 5803 Hz, thus slightly above the target frequency, and the curve for M4 has the maximum at 5072 Hz (indicated by arrow), thus below the target frequency.

frequencies for maskers M3 and M4. The vertical dashed line indicates the target frequency. The curve for M3 (squares) has the maximum at 5803 Hz, thus slightly above the target frequency, and the curve for M4 (circles) has the maximum at 5072 Hz (see arrow), thus below the target frequency. This indicates that off-frequency listening could have provided an advantage for these two maskers. The off-frequency listening advantage was calculated as the difference in signal-to-masker level ratios between the filter centered at the target and the filter corresponding to the maximum. This advantage amounted to 0.5 dB for M3 and 0.2 dB for M4. Thus, excess masking for condition M3M4 may have been overestimated by maximally 0.5 dB.

¹The onset delay is defined as the delay between the onset of the stimulus and the onset of the MOCR-induced gain reduction.

²In some conditions of Bacon *et al.* (1999), the masker-target delay was considerably smaller, but the effect of temporal configuration was not systematically studied.

³The masker frequencies were -55% (M1), -43% (M2), -29% (M3), and $+40\%$ (M4) relative to the target frequency. So, for the lower-frequency maskers all BM measurement studies we are aware of are consistent with linear masker processing. For the higher-frequency masker also the majority of studies suggests linear processing, while one study (Rhode and Recio, 2000) suggests partial compression. A further caveat is that all BM studies are based on measurements in animals and it is unclear how these results can be transferred to humans.

⁴The upper cutoff frequency of the noise was set to the upper edge of the auditory filter (as specified by means of its ERB) centered at the cubic difference tone frequency ($2^*F_M - F_T$, with F_T = target frequency, F_M = masker frequency, and $F_T > F_M$).

⁵As described in Laback *et al.* (2011), the more straight-forward method of varying the masker level often resulted in nonconverging adaptive tracks and was therefore replaced by the iterative approach.

⁶The calculation of excess masking is based on a 3-dB increase of level (corresponding to linear additivity of two equal-level stimuli), multiplied by the growth factor, the result of which is subtracted by 3 dB to account for the linear additivity effect.

⁷We assume that the "corrected" amount of excess masking corresponds to that which would have been obtained if off-frequency listening had been avoided by adding a background noise.

⁸The estimation of suppression was done by predicting the masked threshold for a Gaussian at the frequency of M3 in presence of M3 either with

or without M2 (see also Plack *et al.*, 2002, for the procedure to predict suppression).

⁹While weightings are usually multiplicative, they are included as additive terms because the exponents in Eq. (6) are in dB units.

¹⁰These values correspond to the level increases (in dB) resulting from the summation of n maskers in power units. Note that this formulation assumes equally effective maskers, which is fulfilled in the present study. The dependency of excess masking on k stems from the fact that for the single maskers k has no effect (a equals zero) while the combined masking effect depends on k .

¹¹To check to what extent complex interactions in case of more than two maskers might have affected the model predictions, we also applied a one-stage approach: In this case the values of k_1 , k_2 , k_3 , and k_4 were optimized at once to best predict the results for all masker conditions. The optimum k -values and the corresponding predictions differed only very slightly between the two approaches, thus indicating that the outcomes of the two-stage approach were not importantly affected by effects specific to the multiple-masker conditions.

ANSI (1996). ANSI S3.6, "Specifications for audiometers" (American National Standards Institute, New York).

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (1971). "Properties of 'two-tone inhibition' in primary auditory neurones," *J. Physiol.* **212**, 593–609.

Backus, B. C., and Guinan, J. J. J. (2006). "Time-course of the human medial olivocochlear reflex," *J. Acoust. Soc. Am.* **119**, 2889–2904.

Bacon, S. P., Boden, L. N., Lee, J., and Repovsch, J. L. (1999). "Growth of simultaneous masking for fm fs: Effects of overall frequency and level," *J. Acoust. Soc. Am.* **106**, 341–350.

Balazs, P., Laback, B., Eckel, G., and Deutsch, W. A. (2010). "Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking," *IEEE Trans. Audio, Speech, Lang. Process.* **18**(1), 34–49.

Cheatham, M. A., and Dallos, P. (2001). "Inner hair cell response patterns: implications for low-frequency hearing," *J. Acoust. Soc. Am.* **110**, 2034–2044.

Delgutte, B. (1990). "Two-tone rate suppression in auditory-nerve fibers: dependence on suppressor frequency and level," *Hear. Res.* **49**, 225–246.

Egan, J. P., and Hake, H. W. (1950). "On the masking pattern of a simple auditory stimulus," *J. Acoust. Soc. Am.* **22**, 622–630.

Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.

Greenwood, D. D. (1971). "Aural combination tones and auditory masking," *J. Acoust. Soc. Am.* **50**, 502–543.

Gröchenig, K. (2001). *Foundations of Time-Frequency Analysis* (Birkhäuser, Boston), pp. 1–374.

Humes, L. E., and Jesteadt, W. (1989). "Models of the additivity of masking," *J. Acoust. Soc. Am.* **85**, 1285–1294.

Humes, L. E., Lee, L. W., and Jesteadt, W. (1992). "Two experiments on the spectral boundary conditions for nonlinear additivity of simultaneous masking," *J. Acoust. Soc. Am.* **92**, 2598–2606.

Laback, B., Balazs, P., Necciari, T., Savel, S., Ystad, S., Meunier, S., and Kronland-Martinet, R. (2011). "Additivity of nonsimultaneous masking for short Gaussian-shaped sinusoids," *J. Acoust. Soc. Am.* **129**, 888–897.

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

Lilaonitkul, W., and Guinan, J. J. J. (2009). "Reflex control of the human inner ear: a half-octave offset in medial efferent feedback that is consistent with an efferent role in the control of masking," *J. Neurophysiol.* **101**, 1394–1406.

Lopez-Poveda, E. A., and Alves-Pinto, A. (2008). "A variant temporal-masking-curve method for inferring peripheral auditory compression," *J. Acoust. Soc. Am.* **123**, 1544–1554.

Lopez-Poveda, E. A., and Eustaquio-Martín, A. (2006). "A biophysical model of the inner hair cell: the contribution of potassium currents to peripheral auditory compression," *J. Assoc. Res. Otolaryngol.* **7**, 218–235.

Lutfi, R. A. (1983). "Additivity of simultaneous masking," *J. Acoust. Soc. Am.* **73**, 262–267.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (2001). "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.* **109**, 2852–2861.

Moore, B. C. (1985). "Additivity of simultaneous masking, revisited," *J. Acoust. Soc. Am.* **78**, 488–494.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Emerald, Bingley), pp. 1–441.

- Necciari, T., Balazs, P., Kronland-Martinet, R., Ystad, S., Laback, B., Savel, S., and Meunier, S. (2012). "Perceptual optimization of audio representations based on time-frequency masking data for maximally-compact Stimuli," in *Proceedings of the 45th AES Conference on Applications of Time-Frequency Processing in Audio*, Helsinki, Finland, pp. 103–112.
- Oxenham, A. J., and Moore, B. C. (1995). "Additivity of masking in normally hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.* **98**, 1921–1934.
- Oxenham, A. J., and Plack, C. J. (1998). "Suppression and the upward spread of masking," *J. Acoust. Soc. Am.* **104**, 3500–3510.
- Penner, M. J. (1980). "The coding of intensity and the interaction of forward and backward masking," *J. Acoust. Soc. Am.* **67**, 608–616.
- Plack, C., Oxenham, A., and Drga, V. (2002). "Linear and nonlinear processes in temporal masking," *Acta Acust. Acust.* **88**, 348–358.
- Plack, C. J., and Arifianto, D. (2010). "On- and off-frequency compression estimated using a new version of the additivity of forward masking technique," *J. Acoust. Soc. Am.* **128**, 771–786.
- Plack, C. J., Oxenham, A. J., and Drga, V. (2006). "Masking by inaudible sounds and the linearity of temporal summation," *J. Neurosci.* **26**, 8767–8773.
- Plack, C. J., Oxenham, A. J., Simonson, A. M., O'Hanlon, C. G., Drga, V., and Arifianto, D. (2008). "Estimates of compression at low and high frequencies using masking additivity in normal and impaired ears," *J. Acoust. Soc. Am.* **123**, 4321–4330.
- Recio-Spinoso, A., and Lopez-Poveda, E. (2010). "Basilar membrane responses to simultaneous presentations of white noise and a single tone," in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), pp. 15–23.
- Rhode, W. (1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *J. Acoust. Soc. Am.* **49**, 1218–1231.
- Rhode, W. S., and Cooper, N. P. (1996). "Nonlinear mechanics in the apical turn of the chinchilla cochlea, in vivo," *Aud. Neurosci.* **3**, 101–121.
- Rhode, W. S., and Recio, A. (2000). "Study of mechanical motions in the basal region of the chinchilla cochlea," *J. Acoust. Soc. Am.* **107**, 3317–3332.
- Rhode, W. S., and Recio, A. (2001). "Multicomponent stimulus interactions observed in basilar-membrane vibration in the basal region of the chinchilla cochlea," *J. Acoust. Soc. Am.* **110**, 3140–3154.
- Robles, L., Ruggero, M. A., and Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," *J. Acoust. Soc. Am.* **80**, 1364–1374.
- Rosengard, P. S., Oxenham, A. J., and Braidia, L. D. (2005). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **117**, 3028–3041.
- Ruggero, M. A. (1992). "Responses to sound of the basilar membrane of the mammalian cochlea," *Curr. Opin. Neurobiol.* **2**, 449–456.
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* **101**, 2151–2163.
- Ruggero, M. A., Robles, L., and Rich, N. C. (1992). "Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression," *J. Neurophysiol.* **68**, 1087–1099.
- Russell, I. J., and Murugasu, E. (1997). "Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities," *J. Acoust. Soc. Am.* **102**, 1734–1738.
- Russell, I. J., and Nilsen, K. E. (1997). "The location of the cochlear amplifier: spatial representation of a single tone on the guinea pig basilar membrane," *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2660–2664.
- Sachs, M. B., and Kiang, N. Y. (1968). "Two-tone inhibition in auditory-nerve fibers," *J. Acoust. Soc. Am.* **43**, 1120–1128.
- Yasin, I., and Plack, C. J. (2003). "The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions," *J. Acoust. Soc. Am.* **114**, 322–332.
- Yasin, I., and Plack, C. J. (2005). "The role of suppression in the upward spread of masking," *J. Assoc. Res. Otolaryngol.* **6**, 368–377.
- Yasin, I., and Plack, C. J. (2007). "The effects of low- and high-frequency suppressors on psychophysical estimates of basilar-membrane compression and gain," *J. Acoust. Soc. Am.* **121**, 2832–2841.
- Zwicker, E., and Herla, S. (1975). "Über die Addition von Verdeckungseffekten (On the additivity of masking effects)," *Acustica* **34**, 89–97.