Additivity of nonsimultaneous masking for short Gaussian-shaped sinusoids[a)]

Bernhard Laback[b)] and Peter Balazs
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

Thibaud Necciari, Sophie Savel, Solvi Ystad, Sabine Meunier, and Richard Kronland-Martinet
CNRS Laboratoire de Mécanique et d'Acoustique, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

(Received 19 February 2010; revised 2 November 2010; accepted 2 November 2010)

The additivity of nonsimultaneous masking was studied using Gaussian-shaped tone pulses (referred to as Gaussians) as masker and target stimuli. Combinations of up to four temporally separated Gaussian maskers with an equivalent rectangular bandwidth of 600 Hz and an equivalent rectangular duration of 1.7 ms were tested. Each masker was level-adjusted to produce approximately 8 dB of masking. Excess masking (exceeding linear additivity) was generally stronger than reported in the literature for longer maskers and comparable target levels. A model incorporating a compressive input/output function, followed by a linear summation stage, underestimated excess masking when using an input/output function derived from literature data for longer maskers and comparable target levels. The data could be predicted with a more compressive input/output function. Stronger compression may be explained by assuming that the Gaussian stimuli were too short to evoke the medial olivocochlear reflex (MOCR), whereas for longer maskers tested previously the MOCR caused reduced compression. Overall, the interpretation of the data suggests strong basilar membrane compression for very short stimuli. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3518781]

PACS number(s): 43.66.Dc, 43.66.Ba [CP] Pages: 888–897

I. INTRODUCTION

Auditory masking has been extensively studied for nonsimultaneous (temporal masking) and simultaneous (spectral masking) presentation of masker and target (see Moore, 2003, for a review). Because of the specific demands in the nonsimultaneous and simultaneous masking experiments, the experimental stimuli were almost always broad in the temporal domain, the frequency domain, or both. Quite little is known about nonsimultaneous and simultaneous masking effects for masker and target signals that are well-concentrated in both the time and frequency domains. Such well-concentrated stimuli can be more flexibly arranged in time–frequency space compared to temporally or spectrally broad stimuli while essentially avoiding spectro-temporal overlap. Thus, they appear to be well-suited for studying masking effects with various time–frequency relations between masker and target stimuli. Compared to maskers that are broad in at least one domain, well-concentrated maskers may produce different masking effects, as described below. In the present study, we are concerned with the additivity of masking for multiple well-concentrated maskers that are separated in time.

Due to the uncertainty principle (e.g., Gröchenig, 2001), it is impossible to independently control a signal in both the time and frequency domains. A signal that minimizes this uncertainty as far as possible, i.e., which has minimal spread in the time–frequency domain, is a Gaussian-shaped tone pulse, referred here to as "Gaussian." This elementary signal has a Gaussian shape in both domains. Gaussians have been used in a study attempting to measure the shape of the auditory time–frequency window (van Schijndel *et al.*, 1999) and in studies on the effects of spectral and temporal integration in auditory masking (e.g., van den Brink and Houtgast, 1990).

The present study addresses the question how the masking effects from up to four Gaussian maskers shifted in time relative to a Gaussian target add up. There is a bulk of literature on masking additivity, i.e., the properties according to which the masking effect from one masker adds with the masking effect from another masker (e.g., Humes and Jesteadt, 1989; Oxenham and Moore, 1995; Plack *et al.*, 2006; Plack *et al.*, 2008). Assuming linear additivity of masking in units of intensity, the masked threshold of the 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, it has been found that combining two maskers often results in higher masked thresholds than those predicted by linear additivity. The difference between the linear prediction and the actually measured threshold is referred to as "excess masking." The commonly accepted origin of excess masking for nonsimultaneous maskers, first proposed by Penner (1980), is that the individual maskers are subjected to a compressive nonlinearity before their effects are combined linearly at some higher stage. Consistent with this assumption, different studies showed the amount of excess masking to be large if the individual stimuli (maskers and target) were well separated in time (e.g., Cokely and Humes, 1993).

[a)]Portions of this work were presented at the Acoustics 2008 conference in Paris, 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: Bernhard.Laback@oeaw.ac.at

 © 2011 Acoustical Society of America

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

The special properties of Gaussian maskers described above, most importantly their short duration, could lead to differences in masking additivity compared to previously tested maskers. These differences could hypothetically be caused by differences in basilar membrane (BM) compression for the different types of stimuli. Although the BM starts to be compressive very shortly after the signal onset (Recio *et al.*, 1998), the effective compression for short Gaussian maskers may still differ from that for longer maskers, which in turn may affect the amount of excess masking. Differences in masking additivity could also arise from the properties of the medial olivocochlear reflex (MOCR) which controls the cochlear gain via efferent connections to the outer hair cells (e.g., Backus and Guinan, 2006). This reflex has a delay between the onset of the stimulus and the onset of the gain reduction (referred to as onset delay). This delay could be too long to affect the compressive behavior of the cochlea in response to the target in case of very short stimuli such as Gaussians but not so for longer stimuli. Assuming that the BM is more compressive without activation of the MOCR, it can be hypothesized that excess masking is stronger for a Gaussian masker than for longer maskers. Indeed, the results of some recent nonsimultaneous masking experiments suggest the importance of the masker duration by means of the effect of the MOCR on cochlear compression (Jennings *et al.*, 2009; Wojtczak and Oxenham, 2009; Plack and Arifianto, 2010). One goal of the present study was to measure masking additivity for *very* short Gaussian maskers to rule out the influence of the MOCR and to compare the results with those from the literature for different maskers which might have involved MOCR effects.

Another goal of this study was to explore in a systematic way the dependence of excess masking on the temporal configuration of the maskers and the target. Combinations of up to four forward and backward maskers were tested. This allowed us to assess to what extent the results can be predicted by a model of masking additivity, which incorporates a compressive input/output (I/O) function, reflecting the nonlinear response of the BM, followed by linear summation of the effects of the maskers (Plack *et al.*, 2006; Plack *et al.*, 2008).

In the present study we were mainly interested in the additivity of the "effective" masking produced by multiple maskers, with no particular consideration of the relative contributions of different mechanisms of temporal masking. Nevertheless, it is worthwhile briefly describing some of the potentially underlying mechanisms. Forward masking could be caused by peripheral decay of the response to the masker (Duifhuis, 1973), neural short-term adaptation effects (Smith, 1977), and more central effects of persistence of masker-induced activity (Oxenham, 2001). The origin of backward masking is less understood and may be related to the overlap of the peripheral response to the masker and the target (Duifhuis, 1973).

An important issue in studies of masking additivity is the influence of confounding factors. These confounding factors can lead to overestimation of excess masking either by enhanced target detection in single-masker conditions or by reduced target detection in combined-masker conditions. One such confounding factor in nonsimultaneous masking is so-called "off-time listening" (Robinson and Pollack, 1973; Oxenham and Moore, 1994). In the single-masker conditions, the listener may detect the target at a temporal position remote from the temporal center of the target, where the signal-to-noise ratio (SNR) is higher than at the target position. In the combined-masker conditions with temporally flanking maskers, however, off-time listening provides no SNR gain and may thus cause overestimation of excess masking. To account for this potentially confounding factor, we included also combinations of forward maskers, for which off-time listening does not favor the single-masker conditions.

Describing masking additivity for elementary signals such as Gaussians is potentially interesting with respect to the question how the individual elements, i.e., spectro-temporal components, of more complex sounds contribute to the total masking effect evoked by those sounds. To that end, this study represents a first step within a larger project, studying masking effects for Gaussian maskers with various time–frequency relations relative to a Gaussian target.

II. METHODS

A. Listeners

Five normal-hearing (NH) listeners participated in the experiments. All listeners had absolute thresholds of 15 dB hearing level (HL) or lower at octave frequencies from 125 to 8000 Hz (ANSI, 1996). All except NH24 and NH14 had previous experience in psychoacoustical tasks.

B. Stimuli and apparatus

All masker and target stimuli were Gaussian-shaped sinusoids with a frequency of 4000 Hz, whose waveform and amplitude spectrum are shown in Fig. 1. They were defined by the expression

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

where f_0 is the tone frequency and $\Gamma = \alpha f_0$, α being the shape factor of the Gaussian window. For a given f_0, the shape factor controls the duration and bandwidth of the Gaussian. The equivalent rectangular bandwidth of the Gaussian is given by Γ. The equivalent rectangular duration of the Gaussian equals Γ^{-1}. The value of Γ was chosen based on the study by van Schijndel *et al.* (1999) on intensity discrimination, in which the parameter α of the Gaussian defined above was varied. The goal of the authors was to derive the shape of the "internal" time–frequency observation window of the auditory system. They hypothesized that the intensity discrimination threshold has a maximum at the α-value for which the Gaussian covers the smallest number of time–frequency windows. Interestingly, the maximum was found at the α-value for which the bandwidth of the Gaussian approximates the auditory critical bandwidth. We chose that shape factor ($\alpha = 0.15$) for the Gaussian at $f_0 = 4000$ Hz. Thus, the equivalent rectangular bandwidth of the Gaussians was 600 Hz, and the equivalent rectangular duration was 1.7 ms. The total signal duration, as given by the numeric support, was 9.6 ms. By introducing a phase shift of $\pi/4$, the energy of the signal was

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

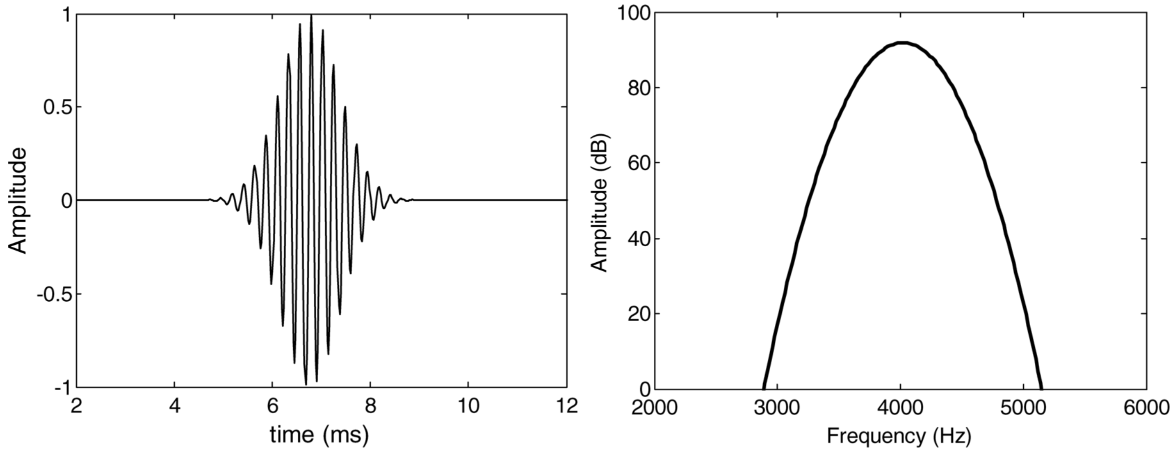

FIG. 1. A Gaussian-shaped tone with f_0 is 4 kHz as defined in Eq. (1). This signal was used both as masker and target. The temporal representation is shown in the left panel and the spectral representation in the right panel.

independent of f_0. This was required for other experiments not reported in this paper, involving different f_0 values. The sound pressure level (SPL) of the Gaussian was specified by measuring the 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, Gentofte, Denmark), passed through an attenuator [PA4; Tucker-Davis Technologies (TDT) Alachua, FL], and a headphone amplifier (HB6, TDT), and routed to the left-ear side of a circumaural headphone (HDA200, Sennheiser). The experiment was performed in a double-walled, sound-attenuated booth.

C. Procedure

Detection thresholds were measured using an adaptive three-interval, forced-choice task (oddity 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. 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 signaled on a computer screen, with a between-interval gap of 200 ms. Within each interval, the first Gaussian was in the temporal center of the interval. Response feedback was provided after each trial by visually highlighting the interval containing the target. In addition, the correctness of the response was indicated by inserting either "correct" or "incorrect" on the screen.

In the adaptive procedure, the target level was varied using the three-down one-up rule that estimates the 79.4% point on the psychometric function (Levitt, 1971). The target level started at a sufficiently high level to be easily audible. 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 8 reversals.

The different experimental conditions were presented in a quasi-randomized order for each listener. Each testing block contained conditions with a fixed number of maskers to avoid switching too often between conditions with potentially different detection cues. A single block lasted about 15 min. After all conditions had been presented once, further repetitions of all the conditions were presented. For each repetition, a new random order of blocks and conditions within the blocks was used. We tracked learning effects for each condition, and continued testing until the following conditions were fulfilled for each condition: (1) At least three valid thresholds were obtained (a threshold was defined as valid if the adaptive staircase converged and if there was no upward or downward trend across the reversals), and (2) there was no upward or downward trend across the last three valid thresholds. The final threshold was determined by averaging over the valid thresholds that showed no apparent upward or downward trend. The mean number of thresholds taken for averaging for each condition and listener was 5.2. Implicit to this design, the first few thresholds for each condition showing learning effects can be conceived as training. The total testing time for one listener was about 6 h.

D. Experimental conditions

The maskers were temporally shifted relative to the target (peak-to-peak distance) by -24 ms (M_1), -16 ms (M_2), -8 ms (M_3), and $+8$ ms (M_4). Thus, there were three forward maskers and one backward masker (see top of Fig. 2). In order to obtain equally effective maskers, each masker had to be adjusted in level. The temporal shifts chosen resulted from pilot tests attempting to find the parameters fulfilling the following requirements: (1) Keep minimum overlap between the maskers and thus use maximum temporal shift, (2) use equal intervals between M_1, M_2, M_3, the target, and M_4, and (3) avoid exceeding a comfortable level for the most distant forward masker (M_1) and the backward masker (M_4), for which the highest masker levels were required. The fact that forward masking is stronger than backward masking resulted in an asymmetric configuration (three forward maskers and one backward masker).

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

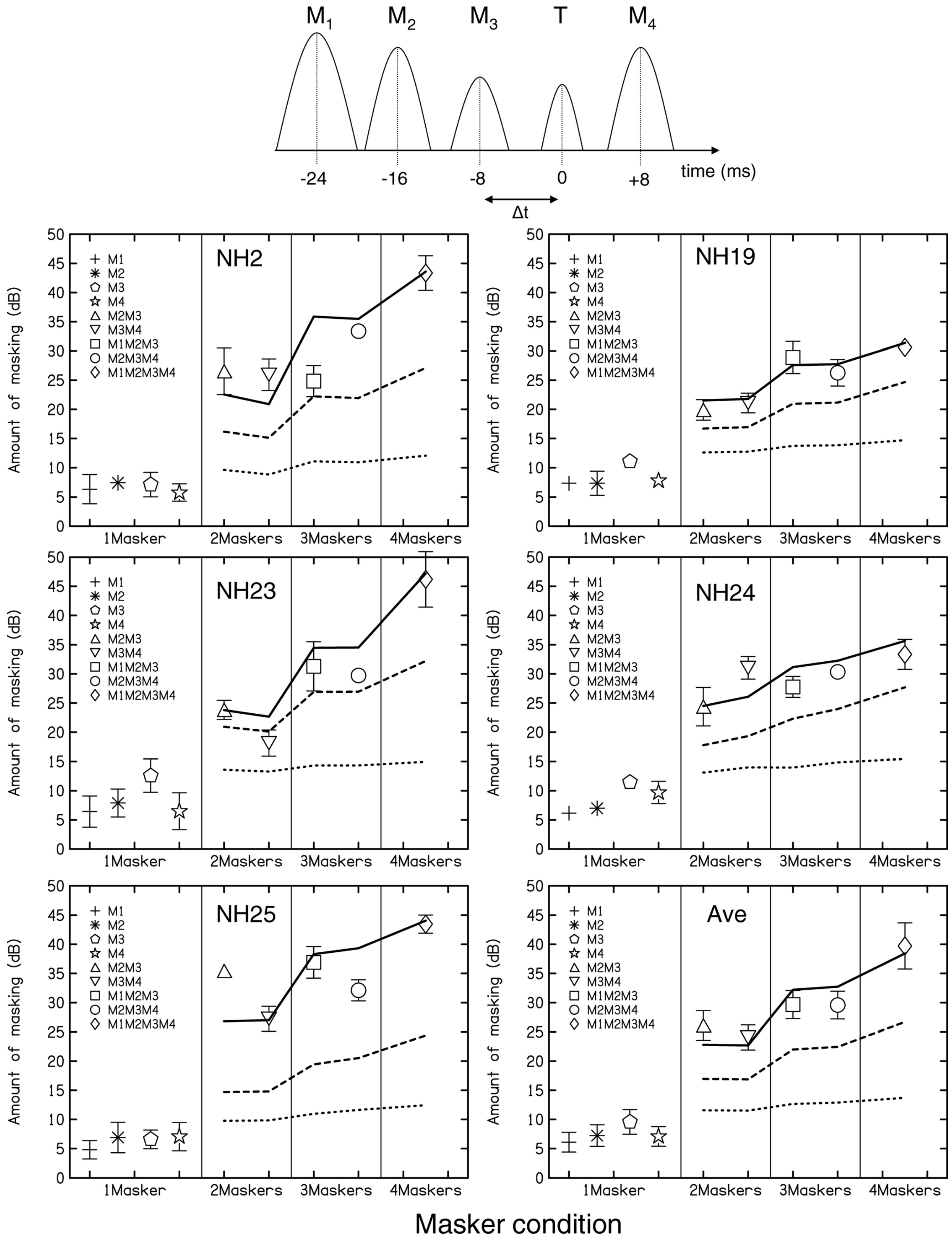

FIG. 2. The open symbols show the experimental results. Each panel provides the data for one listener, except for the bottom-right panel showing the mean results. Different masker combinations are indicated with symbols shown in the legend. Error bars indicate 95% confidence intervals. The dotted lines indicate the predictions from linear masking additivity. The other two lines in each panel show the predictions of the model of masking additivity proposed by Plack *et al*. (2006), using I/O functions best fitting their mean data obtained with long maskers (dashed) and I/O functions best fitting the data of the present study (solid). The small inset at the top of the figure illustrates schematically the temporal arrangement of the stimuli.

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

The experiment consisted of three main stages. In the first stage, the absolute threshold of a single 4000-Hz Gaussian was measured. In the second stage, the level of each masker was assessed which was necessary to produce 8 dB of masking of the target. This was achieved using an iterative approach, where the masker level was adjusted after each run to finally reach the desired masked threshold.[1] On average, about four to six iterations (each representing one masker level) were required. After the final masker level had been determined, further measurements at this level were performed in a separate test session, which represent the masked thresholds for the single-masker conditions. For each listener and masker, the masked threshold had to be significantly above the absolute threshold of the target ($p < 0.05$). Otherwise, the iterative procedure was continued with an adjusted masker level. In the third stage, the main experiment, masked thresholds were measured for selected combinations of maskers. These combinations were M_2M_3, M_3M_4, $M_1M_2M_3$, $M_2M_3M_4$, and $M_1M_2M_3M_4$. It was attempted to select the most interesting combinations while limiting the experiment time. After completion of the main experiment, the single-masker conditions were retested to check for learning effects. No learning was observed, and the data presented below therefore correspond to the mean overall data collected. While the separate testing of single- and combined-masker conditions could theoretically involve order effects, the lack of a systematic change in performance between test and retest for the single-masker condition makes it unlikely that such effects influenced the results significantly.

II. RESULTS AND INTERIM DISCUSSION

A. Single maskers

Table I presents individual absolute thresholds of the target (see first column) and the masker levels necessary to produce about equal amounts of masking of the target (i.e., masker levels for which the target level at threshold was about 8 dB above absolute threshold; see columns 2–5). As expected from forward masking studies, the equally effective masker level increased with increasing temporal delay between the masker and the target.

Figure 2 presents the results of the individual listeners and the mean results in separate panels. The leftmost column of each panel shows the amount of masking for the single maskers. Masker M_1, presented 24 ms before the target at a sensation level (SL) of 59 dB, produced 6.1 dB of masking on average across listeners. This result is in good agreement with data collected by Necciari *et al.* (2008), using identical stimuli and procedures but different subjects and equipment (amount of masking for a masker at 60 dB SL = 5 dB), and with the results from Widin and Viemeister (1980), obtained with short 1000-Hz tone bursts (amount of masking for a masker of comparable level presented 13 ms before the target = 8.7 dB). Masker M_3 (presented 8 ms before the target) produced 9.6 dB of masking, which is again consistent with the study by Widin and Viemeister (amount of masking for a masker presented 6.5 ms before the target = 10 dB). The masker levels required to cause about the same amount of masking of the target are similar for M_4 (+8 ms) and M_1 (−24 ms), despite the largely different temporal separations, indicating much weaker backward masking than forward masking. This is consistent with previous studies (e.g., Wilson and Carhart, 1971; Penner, 1980; Oxenham and Moore, 1994).

B. Multiple maskers

The open symbols in columns 2–4 of each panel in Fig. 2 show the amount of masking for all combinations of maskers tested. Despite considerable across-listener variability, we focus below on the mean data, supported by a statistical analysis. Specifically, a repeated-measures analysis of variance (ANOVA) was performed with the factor group, which separated the data into conditions with one, two, three, or four maskers. The main effect of the factor group was highly significant ($F_{3,190} = 512$, $p < 0.0001$). The differences between the factor levels were analyzed with a Tukey's HSD (honestly significant difference) *post-hoc* test.

The amount of masking was 17.6 dB higher with two maskers than with a single masker ($p < 0.0001$), on average across conditions. Masking increased further by 4.6 dB with three maskers ($p < 0.0001$) and increased even further by 10.1 dB with four maskers ($p < 0.0001$). The total increase in masking from one to four maskers was 32.3 dB. There

TABLE I. Absolute thresholds for the 4000-Hz target and masker levels resulting in masked thresholds of the target approximately 8 dB above its absolute threshold, for each of the four maskers.

Subject	Absolute threshold of target (dB SPL)	Equally effective masker level (dB SPL)			
		M_1 (−24 ms)	M_2 (−16 ms)	M_3 (−8 ms)	M_4 (+8 ms)
NH2	27.3	85.3	84.3	73.3	86.3
NH19	19.3	79.3	70.3	60.3	80.3
NH23	29.3	87.3	83.3	71.3	80.3
NH24	23.3	79.3	70.3	63.3	77.3
NH25	24.3	87.3	73.3	63.3	79.3
Mean	24.7	83.7	76.3	66.3	80.7
SD[a]	3.8	4.1	7.0	5.7	3.4

[a]Standard deviation.

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

was no statistical difference in the additivity of masking between combining forward maskers only and combining forward maskers with the backward masker (paired t-tests with Bonferroni correction: $p \geq 0.48$). However, exceptions were noted in individual results (e.g., NH2 showed more masking in condition $M_2M_3M_4$ than in condition $M_1M_2M_3$).

The dotted lines in Fig. 2 show the amount of masking corresponding to a linear additivity model, calculated by linear summation (in intensity units) of the amounts of masking produced by each individual masker. According to this model, combining exactly equally effective maskers increases the amount of masking relative to a single masker by 3 dB for two maskers, 4.8 dB for three maskers, and 6 dB for four maskers. It is obvious that the data show a large amount of excess masking. On average, excess masking was 13.5 dB in the two-masker conditions, 16.9 dB in the three-masker conditions, and 26.0 dB in the four-masker condition.

The results can be compared to the quite extensive literature on the additivity of nonsimultaneous masking. In most of the studies the target was a short tone burst, similar to our Gaussian. The maskers were bandpass-filtered or broadband noise, sinusoids, or filtered clicks. In most but not all studies, the individual maskers were adjusted in level so that they were equally effective in masking the target. The results are compared in terms of the SL of the target. To our knowledge, only one study investigated the additivity of nonsimultaneous masking for up to four maskers (Penner, 1980). The level of the target for the single maskers was fixed to 8 dB SL, as in our study. The maskers were wide-band noises, and the target was a bandpass-filtered click. The amount of excess masking was 5 dB for two maskers and 14 dB for four maskers, thus 8.5 and 12 dB, respectively, below our results. Penner (1980) and Penner and Shiffrin (1980) also tested the additivity of masking for pairs of equally effective maskers at different target SLs, using the same types of stimuli. For high target SLs excess masking increased up to 12 dB. For a target SL of about 28 dB, the excess masking reported by Penner (1980) was comparable to that observed in our study for a target SL of 8 dB. Patterson (1971) tested in three frequency regions (500, 1500, and 2500 Hz) and at a mean SL of 10.2 dB. The amount of excess masking was about 11 dB, on average across the three signal frequencies. Wilson and Carhart (1971) found about 7 dB of excess masking at target SLs of about 20 dB.

Widin and Viemeister (1980) found about 8 dB of excess masking for low target SLs (approximately comparable to our study). Oxenham and Moore (1995) reported quite low excess masking (of the order of 3 dB) for SLs comparable to our study and excess masking of up to about 10 dB for the highest target SL tested (25 dB). Oxenham and Moore (1994) observed little excess masking for target SLs comparable to our study but up to about 8–15 dB for higher levels. Cokely and Humes (1993) tested only at high target SLs (>20–30 dB) and found excess masking of up to 18 dB. Plack and O'Hanlon (2003), Plack *et al*. (2006), and Plack *et al*. (2008) reported little excess masking at target SLs comparable to our study but up to about 15 dB excess masking at higher levels. Finally, the most recent study by Plack and Arifianto (2010) found excess masking between about 9 and 14 dB (depending on the test condition) at the lowest target level (10 dB SL) and between 19 and 25 dB (depending on the test condition) at 30 dB SL.

In summary, the amount of excess masking we observed is markedly larger than that reported in most of the studies of masking additivity involving comparable target levels. For higher target levels, however, those studies show results similar to ours. This raises the question of which aspect of the experiment caused the larger excess masking in our study.

A first possible reason is that we coincidentally sampled listeners that behave differently from those of the previous studies. This explanation is very unlikely, because the difference in excess masking between the pooled data from the previous studies and the present study is significant, close to the 1% level (Welch's t-test: $p = 0.027$).

A second possible explanation could be that the more "complex" nature of the combined-masker stimuli made it more difficult for the listeners to detect the target in those conditions than in the single-masker conditions. Masking effects, which cannot be attributed to peripheral overlap of the stimuli and hence involve more central factors, are often referred to as informational masking (e.g., Watson, 1987). For example, higher degrees of target–masker similarity can result in higher masked thresholds (e.g., Durlach *et al*., 2003). This could be relevant in our experiment, because masker and target stimuli were identical. However, it should be considered that we used a three-interval, forced-choice task where the listeners did not need to identify the target but instead could have used any difference cue to identify the interval containing the target. Therefore, if we assume that the listeners based their decision on the comparison across the three intervals, we consider it unlikely that the target–masker similarity contributed importantly to excess masking. Furthermore, the degree of uncertainty about the masker and target signals presented in a given experimental trial can influence the amount of masking (e.g., Durlach, 2006). A certain amount of masking not attributable to peripheral overlap of masker and target has been reported even under conditions of minimal stimulus uncertainty (e.g., Leibold and Werner, 2006). Although we attempted to minimize the amount of uncertainty by using an adaptive procedure with a fixed masker within a run, some uncertainty might still have been present, and this uncertainty might have increased with the number of maskers involved.

A third possible explanation for the discrepancy between the present and past results is differences in stimulus properties. While our target stimulus was quite comparable to the target stimuli used in the literature, the maskers were different. We used short tone pulses, whereas most studies used noise or tone maskers with comparatively long durations. Only Penner (1980, experiment 1), Patterson (1971), and one condition in Plack and Arifianto (2010) involved short maskers. The latter two studies found 11 and 14 dB of excess masking, respectively, which are the largest values from all studies and are within the range of the results from our study. The comparison across studies in terms of masker duration points to the role of masker duration. In fact, assuming static compression in the cochlea, the maskers themselves should not influence the amount of excess masking for equally

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

effective maskers. It is rather the signal level at masked threshold along the compressive function of the cochlea that should matter (appendix of Oxenham and Moore, 1995).

A fourth potential explanation for the greater amount of excess masking in our study is the effect of the MOCR. The MOCR controls the cochlear gain via efferent connections to the outer hair cells. A change in gain could affect the compression of the target and thus the amount of excess masking. The MOCR has an onset delay of about 25 ms (Backus and Guinan, 2006). Wojtczak and Oxenham (2009) reported that the differences they found in forward masking between long and short maskers can be explained by a model incorporating MOCR-controlled gain reduction. Jennings *et al.* (2009) found some indications that the presence of a precursor stimulus preceding a 20-ms forward masker reduces the cochlear gain. With regard to masking additivity, the data from the literature using long-duration maskers could have been affected by MOCR gain reduction. Conversely, in the present study, the gain reduction is unlikely to have occurred because the intervals between the onset of the first masker and the target (8–24 ms) were all shorter than the average MOCR onset delay (Backus and Guinan, 2006).[2] Studies of masking additivity are largely consistent with this explanation: They mostly reported a lower amount of excess masking compared to our study when the interval between the onset of the first masker and the target exceeded about 25 ms. The recent study by Plack and Arifianto (2010) involved pairs of forward maskers with an interval from the onset of the first masker to the onset of the target of either 210 or 30 ms. The mean slope of the I/O functions estimated from the data was reported to be significantly shallower for the shorter interval. This result is consistent with the idea that the condition with the larger interval involved the activation of the MOCR, causing less BM compression (and thus a steeper I/O function) compared to the condition with the short interval. The effect of the interval between the onset of the first masker and the target was relatively small, which might indicate some influence of the MOCR even for the condition with the shorter (30-ms) interval, given that it slightly exceeded the mean onset delay of the MOCR. Another interesting observation can be made from the data of Cokely and Humes (1993), who reported increasing excess masking when increasing the gap between the offset of the second masker and the onset of the target from 10 to 30 ms. This result might reflect the decay of the MOCR following the offset of the masker and thus the recovery of BM compression.

It has been suggested that excess masking for combinations of a forward and a backward masker may arise from off-time listening effects. However, both past studies (Cokely and Humes, 1993; Penner, 1980) and the present study do not support this off-time listening explanation because they reported similar amounts of masking additivity regardless of whether forward maskers only were used or whether forward and backward maskers were combined.

C. Modeling

The thresholds for the combined maskers were modeled using a model of the additivity of masking described in Plack *et al.* (2006). With this model they were able to predict target thresholds for the combination of two temporally non-overlapping maskers based on the target thresholds for each of the single maskers. The model, based on the work of Penner (1980), assumes that excess masking is a result of the maskers being independently subjected to a compressive I/O function before their effects are combined within a linear temporal integrator. This model assumption has been adopted in different forms in the literature (e.g., Humes and Jesteadt, 1989; Oxenham and Moore, 1994; Plack and O'Hanlon, 2003). Plack *et al.* (2006) modeled the I/O function by a third-order polynomial function which is defined here in units of decibels:

$$f(x) = ax^3 + bx^2 + cx, \tag{2}$$

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). The model assumes that the detection of the target depends on the target-to-masker ratio after compression and summation and that this ratio is constant at threshold for all conditions. Note that the model does not process the stimuli themselves but rather uses the masked thresholds for the individual maskers as model input. A measure of the masking effect, E, can be taken as the target intensity at masked threshold after compression,

$$E = f(S), \tag{3}$$

where S is the target intensity at threshold. Assuming that the effects of the maskers $(M_1, M_2, \ldots, M_n)$ add linearly,

$$E_{\mathrm{COMB}} = E_{M_1} + E_{M_2} + \cdots + E_{M_n}, \tag{4}$$

where E_{M_1}, E_{M_2}, and E_{M_n} are the masking effects of each masker separately and E_{COMB} is the combined masking effect of all maskers. Substituting from Eq. (3) and solving for S gives

$$S_{\mathrm{COMB}} = f^{-1}\left(10 \cdot \log\left(10^{\frac{f(S_{M_1})}{10}} + 10^{\frac{f(S_{M_2})}{10}} + \cdots + 10^{\frac{f(S_{M_n})}{10}}\right)\right) \tag{5}$$

where S_{M_1}, S_{M_2}, and S_{M_n} are the target intensities at threshold in the presence of the individual maskers and S_{COMB} is the target intensity at threshold in the presence of all n maskers combined. Using this equation in combination with Eq. (2) as the function f the thresholds for the single maskers $(S_{M_1}, S_{M_2}, \ldots, S_{M_n})$ were used as the input of the model, and the thresholds in the presence of the combined maskers (S_{COMB}) were predicted. Plack *et al.* derived the coefficients of the polynomial best fitting their data of masking additivity for pairs of forward maskers. For their mean data, the best fitting coefficients were $a = 4.3 \times 10^{-5}$, $b = -9.9 \times 10^{-3}$, and $c = 0.913$.[3] These coefficients were applied to the present data, and the corresponding predictions are shown in Fig. 2 with dashed lines.[4] It is obvious that the model underestimates the amount of excess masking for all masker combinations. This underestimation was expected because our maskers produced more excess masking than those used in Plack *et al.* (2006).

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

In the next step, we attempted to find the shape of the I/O functions required to best predict our data. Therefore, for each listener, the coefficients of the polynomial in Eq. (2) (a, b, and c) were determined to minimize the sum of the squared deviations of the model predictions from the thresholds in the combined-masker conditions. The differential of the function f was not allowed to fall below 0 or exceed 1 over the range of thresholds measured for each listener. The masked threshold predictions derived from the best fitting I/O functions for the individual listeners are shown in Fig. 2 with solid lines. The model provides a reasonable fit to the data, except for some individual data points in listeners NH2 and NH25, which deviate from the pattern in the mean data. The upper left panel of Fig. 3 shows the best fitting I/O functions for the individual listeners. The upper right panel shows the function best fitting our mean data together with the function derived by Plack *et al*. (2006) for their mean data, the latter being similar to physiological measurements in the chinchilla cochlea. The I/O functions in Fig. 3 are arbitrarily normalized to give an 80-dB output for a 50-dB input. Within the range of levels tested in the present study and in Plack *et al*. (2006), the mean function derived from the present study is clearly shallower, thus more compressive. The amount of inter-individual variability in the shapes of the I/O functions seems, to be comparable in the present study and in previous studies using the same model (Plack *et al*., 2006; Plack *et al*., 2008; Plack and Arifianto, 2010).

The lower panels of Fig. 3 show the slopes, i.e., the derivatives, of the I/O functions shown in the upper panels. They correspond to the compression exponent value for any given input level. At low levels (up to at least 40 dB SPL), the mean slope is systematically shallower in our study than in the study by Plack *et al*. (2006). Note that the apparent inter-individual variability in the I/O function slopes seems to be comparable to that observed in the previous studies using the same model. Overall, the modeling results illustrate that our experimental data for short forward maskers can be explained by assuming more compressive cochlear I/O functions than for longer forward maskers tested in the literature. As already mentioned, this difference could be explained by the MOCR reflex, which might have reduced the cochlear gain in experiments using long forward maskers and was most likely absent in our experiment with short maskers.

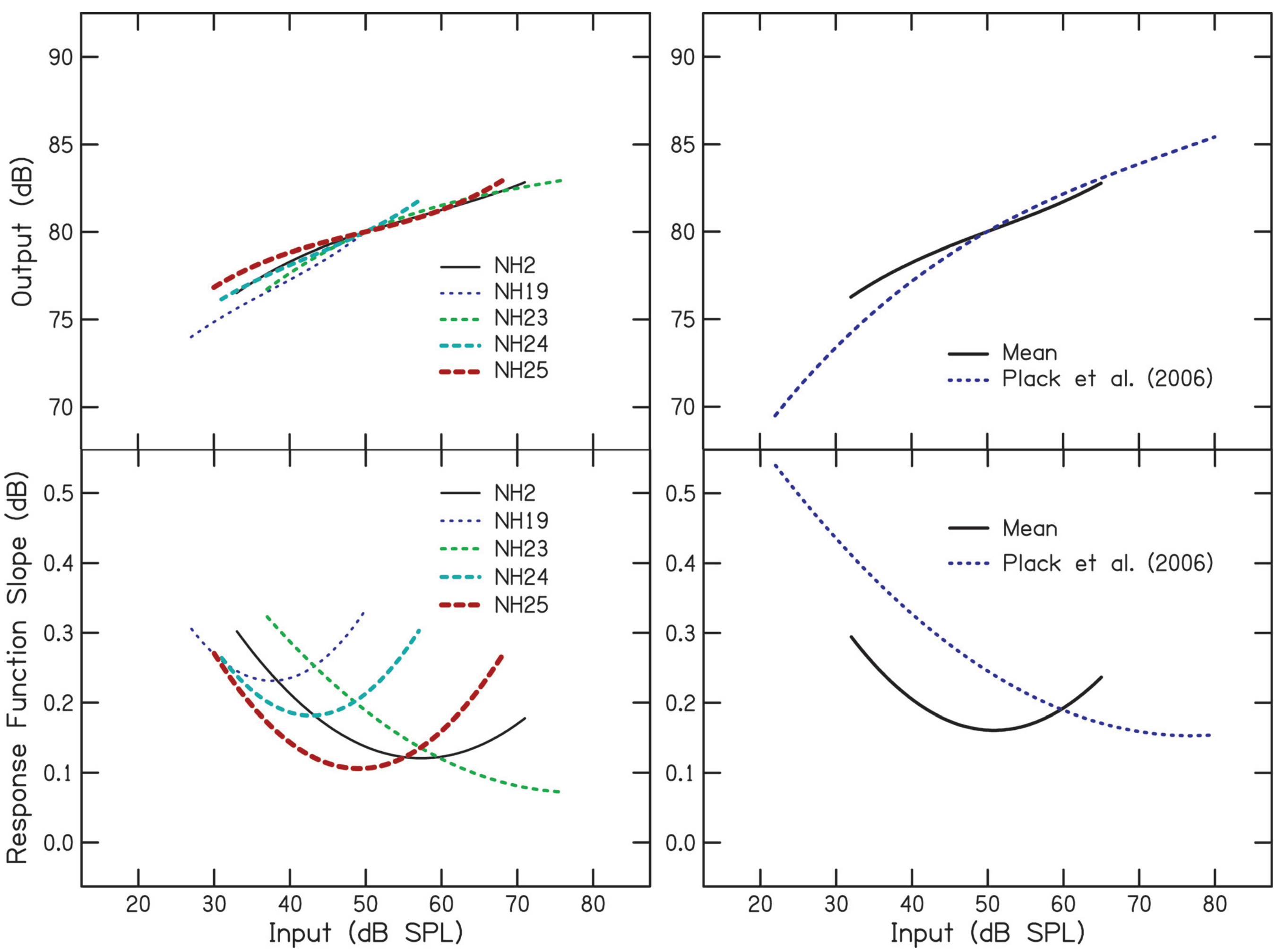

FIG. 3. (Color online) The upper panels show the I/O functions derived from the experimental results of the present study, using the masking additivity model. The left panel shows the results for the individual listeners. The right panel compares the function derived from the mean data of the present study ($a = 1.2 \times 10^{-4}$, $b = -0.0186$, and $c = 1.1217$) with the function derived from the mean data from Plack *et al*. (2006) using long maskers. The lower panels show the slopes, i.e., derivatives of the I/O functions from the upper panels.

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

III. GENERAL DISCUSSION AND CONCLUSIONS

We studied the additivity of nonsimultaneous masking using short Gaussian-shaped sinusoids both as maskers and target. Up to four maskers were tested which were temporally separated from the target. An important finding is that combinations of Gaussian-shaped tone pulses reveal rather strong amounts of excess masking. The amount of excess masking was considerably larger than that observed in most previous studies using longer maskers and comparable target levels. In an attempt to explain this difference, we identified a potential origin, the MOCR, which controls the cochlear gain via efferent connections to the outer hair cells. We suggested that the MOCR reduces the compression of the BM for forward masker configurations where the interval between the onset of the first masker and the target exceeds the MOCR onset delay (25 ms) but not for the short Gaussian maskers used in the present study (interval between onset of first masker and target: 8–24 ms). According to this explanation, the lack of MOCR effect for the Gaussians results in a relatively stronger compression of the target. For longer maskers that are supposed to evoke the MOCR, such a compression rate appears to be reached only at higher levels. Consistent with the experimental results, a model of masking additivity was shown to systematically underestimate the amount of excess masking when using parameters for the I/O function that best fitted previous data of masking additivity obtained with long forward maskers. The I/O functions best fitting our data were more compressive than those best fitting the data for long maskers.

The explanation for stronger excess masking with short maskers based on the MOCR seems at first sight to be inconsistent with the conclusion from modeling work of Jennings *et al*. (2009), who suggested that the MOCR may actually *cause* excess masking in certain configurations, namely when the second masker had a frequency differing from that of the target. However, for the condition in Jennings *et al*. where both maskers had the same frequency as the target, as in our study, a model of masking additivity predicted the data equally accurately. This is consistent with our conclusion. Our results are also consistent with the recent study by Plack and Arifianto (2010), who reported shallower I/O functions (and thus more excess masking) in some cases when reducing the interval from the onset of the first masker to the target. The authors suggested a possible influence of the MOCR.

Using up to four maskers allowed us to test whether the model of masking additivity was able to predict masking additivity for more than two maskers. To our knowledge, this model has so far only been tested for two maskers. The experimental data presented here showed a monotonic increase in excess masking with increasing number of maskers. When all four maskers were combined, the mean amount of excess masking was as high as 26 dB. The model was able to predict these data relatively well. This indicates that the model might be applicable for arbitrary numbers of maskers, as occurring for real-world sounds.

It is interesting to consider the nonlinear masking additivity for a signal with an equivalent rectangular duration of only 1.7 ms. Nonlinear additivity of nonsimultaneous masking is thought to be a result of individual maskers being independently compressed on the BM and subsequently being summed linearly. Our finding of highly nonlinear additivity indicates that the Gaussians are subject to strong BM compression. This is consistent with the physiological finding that the BM starts to be highly compressive within 500 and 700 μs after the onset of a click (Recio *et al*., 1998).[5]

A factor potentially contributing to the nonlinear additivity of masking is the increasing complexity of the background sound, i.e., the masker against which the target has to be detected. According to this explanation, excess masking would be a natural consequence of adding more information to the masker, involving some type of informational masking (e.g., Watson, 1987). While our experimental design probably minimized the influence of effects such as masker–target similarity or uncertainty about the masker or target, we do not know the actual contribution of such effects, neither in our experiment nor in experiments on masking additivity in general (see Durlach, 2006). Even though our results are consistent with the physiological observation of strong BM compression shortly after the signal onset, future studies should explore the potential influence of informational masking. It could be tested if excess masking decreases to some extent when enhancing the salience of the target, for example, by presenting the stimuli binaurally and introducing a large interaural time difference between masker and target. Note, however, that the potential involvement of informational masking is a general issue which could be relevant for many studies on masking additivity.

Using temporally and spectrally narrow stimuli such as Gaussians has the advantage over temporally and/or spectrally broad stimuli that they can be arranged flexibly with respect to time and frequency. Using such temporally and spectrally narrow masker and target stimuli may allow investigating time–frequency masking effects involving short time constants. Studies currently underway investigate the additivity of masking for multiple Gaussian maskers with various time–frequency relations (Laback *et al*., 2008). Combining the data from those studies with a recent study on the spread of time–frequency masking of a single Gaussian masker (Necciari *et al*., 2008) and with the present study may serve as a basis for the development of a model predicting time–frequency masking effects between the individual elements, i.e., spectro-temporal components, of real-world sounds. Every real-world sound can indeed be decomposed into a time–frequency matrix of Gaussian components with appropriate amplitudes and phases (Gröchenig, 2001), so-called "atoms" which are well-concentrated in the time–frequency domain. Having a model for time–frequency masking effects between the atoms may then allow predicting the perceptual contribution of each component of the matrix, i.e., if it can be removed without audible difference. This would represent an extension of the concept of the irrelevance filter model (Balazs *et al*., 2010), incorporating time–frequency masking effects. Perceptual audio codecs such as MP3 also decompose sounds into a matrix of time frames and frequency bands, which can be considered as atoms. Describing the masking effect produced by each of these atoms and their additivity might contribute to the improvement of the efficiency of such codecs. However, it is

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

unclear how successful such approaches can be, given the high degree of complexity and nonlinearity of the auditory system.

ACKNOWLEDGMENTS

We would like to thank the associate editor, Chris Plack, and two anonymous reviewers for helpful comments on earlier versions of the manuscript. We thank Gwenael Toupin for programming the software routines required for the experiments and for running the majority of experiments. We thank Katharina Egger for collecting and preparing some of the data and Piotr Majdak for supporting the development of the software for the experiments. We also thank Dr. Larry Humes and Dr. Elisabeth Strickland for stimulating suggestions. The work was partly supported by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) (project MA07-025), the Wissenschaftlich-Technologische Zusammenarbeit (WTZ) (project Amadee), and the Agence Nationale de la Recherche (ANR) (project "SenSons").

[1]Note that a more direct approach to determine equally effective masker levels would be to vary the masker level instead of the target level (e.g., Oxenham and Moore, 1995; Plack and O'Hanlon, 2003). However, when we used this method in pilot tests we often obtained non-converging adaptive tracks. Therefore, we discarded this method. The difficulty with this method might result from the fact that the level of the masker changes from trial to trial, which could make it harder to build an internal reference against which the presence of the target is detected. Such difficulties were not reported in studies using longer maskers (e.g., Oxenham and Moore, 1995).

[2]Note that the MOCR onset delays reported in Backus and Guinan (2006) for individual listeners range from 15 to 40 ms, with a mean value of 25 ms. Since we found strong excess masking even for condition M_2M_3, for which the interval from the onset of the first masker to the onset of the target amounted to only 16 ms, it is unlikely that the MOCR was involved.

[3]These coefficients, best fitting the mean data across subjects in Plack *et al*. (2006), were reported in Jennings *et al*. (2009).

[4]One caveat here is that Plack *et al*. (2006) reported only model predictions for subject-specifically fit polynomials whereas we applied here the polynomial best fitting their mean data. However, the polynomial best fitting their mean data appears to be representative of their individually fit polynomials with respect to the shape, suggesting that the predictions would not be systematically different.

[5]The measurements were made in the 8–10 kHz region of the BM of the chinchilla.

ANSI. (**1996**). "Specifications for audiometers," Technical Report S3.6, ANSI, New York.

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

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**, 34–49.

Cokely, C. G., and Humes, L. (**1993**). "Two experiments on the temporal boundaries for the nonlinear additivity of masking," J. Acoust. Soc. Am. **94**, 2553–2559.

Duifhuis, H. (**1973**). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. **54**, 1471–1488.

Durlach, N. I., Mason, C. R., Kidd, G. J., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Durlach, N. (**2006**). "Auditory masking: Need for improved conceptual structure," J. Acoust. Soc. Am. **120**, 1787–1790.

Gröchenig, K. (**2001**). *Foundations of Time-Frequency Analysis* (Birkhäuser, Boston), pp. 1–361.

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

Jennings, S. G., Strickland, E. A., and Heinz, M. G. (**2009**). "Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking," J. Acoust. Soc. Am. **125**, 2172–2181.

Laback, B., Balazs, P., Toupin, G., Necciari, T., Savel, S., Meunier, S., Ystad, S., and Kronland-Martinet, R. (**2008**). "Additivity of auditory masking using Gaussian-shaped tones," in *Proceedings of Acoustics'08*, Paris, pp. 3889–3894.

Leibold, L. J., and Werner, L. A. (**2006**). "Effect of masker-frequency variability on the detection performance of infants and adults," J. Acoust. Soc. Am. **119**, 3960–3970.

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

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, London, United Kingdom), pp. 1–413.

Necciari, T., Savel, S., Meunier, S., Ystad, S., Kronland-Martinet, R., Laback, B., and Balazs, P. (**2008**). "Auditory masking using Gaussian-windowed stimuli," in *Proceedings of Acoustics'08*, Paris.

Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.

Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.

Oxenham, A. J. (**2001**). "Forward masking: Adaptation or integration?" J. Acoust. Soc. Am. **109**, 732–741.

Patterson, J. H. (**1971**). "Additivity of forward and backward masking as a function of signal frequency," J. Acoust. Soc. Am. **50**, 1123–1125.

Penner M. J. (**1980**). "The coding of intensity and the interaction of forward and backward masking," J. Acoust. Soc. Am. **67**, 608–616.

Penner, M. J., and Shiffrin, R. M. (**1980**). "Nonlinearities in the coding of intensity within the context of a temporal summation model," J. Acoust. Soc. Am. **67**, 617–627.

Plack, C. J., and O'Hanlon, C. G. (**2003**). "Forward masking additivity and auditory compression at low and high frequencies," J. Assoc. Res. Otolaryngol. **4**, 405–415.

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.

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.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.

Robinson, C. E., and Pollack, I. (**1973**). "Interaction between forward and backward masking: A measure of the integrating period of the auditory system," J. Acoust. Soc. Am. **53**, 1313–1316.

Smith, R. L. (**1977**). "Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects," J. Neurophysiol. **40**, 1098–1111.

van den Brink, W. A., and Houtgast, T. (**1990**). "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window," J. Acoust. Soc. Am. **105**, 3425–3435.

Watson, C. S. (**1987**). "Uncertainty, informational masking and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.

Widin, G. P., and Viemeister, N. F. (**1980**). "Masker interaction in pure-tone forward masking," J. Acoust. Soc. Am. **68**, 475–479.

Wilson, R. H., and Carhart, R. (**1971**). "Forward and backward masking: Interactions and additive," J. Acoust. Soc. Am. **49**, 1254–1263.

Wojtczak, M., and Oxenham, A. J. (**2009**). "On- and off-frequency forward masking by Schroeder-phase complexes," J. Assoc. Res. Otolaryngol. **10**, 595–607.

Downloaded 04 Apr 2011 to 194.199.99.186. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp