Hearing Research 420 (2022) 108514

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Temporal integration contributes to the masking release by amplitude modulation[1]

Hisaaki Tabuchi[a,*], Bernhard Laback[b]

[a] Department of Psychology, University of Innsbruck, Innrain 52f, 6020 Innsbruck, Austria
[b] Austrian Academy of Sciences, Acoustics Research Institute, Wohllebengasse 12-14, 1040 Vienna, Austria

ARTICLE INFO

Article history:
Received 24 June 2021
Revised 29 March 2022
Accepted 4 May 2022
Available online 12 May 2022

Keywords:
Temporal integration
Dip-listening integration
Multiple-looks integration
Envelope fluctuation
Medial olivocochlear (MOC) system
Simultaneous masking
Schroeder-phase harmonic complex
Compression
Physiology-based auditory model

ABSTRACT

Modulated maskers produce less amount of masking than unmodulated maskers, an effect referred to as masking release (MR). Both listening in the temporal dips and fast cochlear compression have been suggested as underlying mechanisms. We addressed the role of dip listening by measuring temporal integration in simultaneous masking using Schroeder-phase harmonic complexes (SPHC) with various phase curvatures. In an experiment with six normal-hearing listeners, SPHC masker and pure-tone target stimuli were covaried in duration at a high masker level. The MR increased with stimulus duration, suggesting integration of target information across multiple masker dips. The duration dependence of the MR was predicted by a physiology-inspired model based on the temporal envelope modulation strength in the auditory periphery. The modeling analysis suggested that listeners detect the presence of the target by a reduction in fluctuation strength that results primarily from a decline of F0-based response peaks, an effect known as synchrony capture. The detailed pattern of masked thresholds across various masker phase curvatures was not predicted by the model, suggesting that its phase response does not well fit the human phase response. Overall, temporal integration across neural envelope features associated with the masker dips seems to contribute to the MR with SPHCs.

1. Introduction

Simultaneous masking refers to the reduced audibility of a target signal in presence of a simultaneous masker. It is well established that imposing amplitude modulation on a masker increases the target audibility relative to a masker without amplitude modulation (e.g., Buus, 1985). The difference of masked thresholds between modulated and unmodulated masker conditions is often referred to as masking release (MR). An elegant way of demonstrating MR is to use Schroeder-phase harmonic complexes (SPHC; Mehrgardt and Schroeder, 1983; Smith et al., 1986) which allow to vary the stimulus envelope fluctuation (i.e., modulation depth) by varying a "phase" parameter while keeping the power spectrum constant. One explanation for the MR is so-called dip listening, i.e., the ability to detect a target in the temporal dips of the masker stimulus, where the target-to-masker energy ratio (analogous to signal-to-noise ratio, SNR) is high (Mehrgardt and Schroeder, 1983; Kohlrausch and Sander, 1995). More generally, the dip listening advantage has also been studied in tone detection with amplitude modulated or unmodulated noise maskers (Bacon et al., 1997; Bacon and Lee, 1997; Gleich et al., 2007), with co-modulated or non-comodulated noise maskers (Schooneveldt and Moore, 1989; Buss et al., 2012), or in the detection of a short tone at different temporal positions within the cycle of an amplitude-modulated tone masker or SPHC masker, a method referred to as masking period pattern (Kohlrausch and Sander, 1995; Summers, 2000; Wojtczak et al., 2001). Dip listening has also been studied with respect to speech recognition using SPHC maskers (Summers and Leek, 1998; Green and Rosen, 2013; Deroche et al., 2013) or other masker types (Bacon et al., 1998; Freyman et al., 2012; Shen and Pearson, 2017).

A second explanation for MR, which is not mutually exclusive with dip listening, involves the importance of fast-acting compression, primarily due to outer-hair-cell activity in the cochlea (Oxenham and Dau, 2001a, 2001b, 2004; Alcántara et al., 2003). This compression explanation has been supported by the finding of strong MR even in forward masking (Carlyon and Datta, 1997; Wojtczak and Oxenham, 2009) where dip listening is not possible. Assuming that the target power spectrum is integrated at least over one modulation period of the masker, and that the integration takes place after the fast-acting compression, a "modulated"

[1] Portions of this work were presented at the 43rd MidWinter Meeting of the Association for Research in Otolaryngology, San Jose, 2020.

* Corresponding author.
E-mail address: Hisaaki.Tabuchi@uibk.ac.at (H. Tabuchi).

https://doi.org/10.1016/j.heares.2022.108514

masker causes a smaller integrated excitation than an "unmodulated" masker, resulting in less masking (Carlyon and Datta, 1997). Another line of evidence for the importance of compression in the MR comes from listeners with cochlear hearing loss, i.e., reduced cochlear compression, who typically show reduced MR (Summers and Leek, 1998; Summers, 2000; Oxenham and Dau, 2004).

A recent study (Tabuchi et al., 2016) revisited the role of fast-acting compression in simultaneous masking by testing the effect of a precursor presented immediately before the masker-plus-target stimulus, intended to reduce fast-acting compression by means of activation of the efferent system, particularly the medial-olivocochlear (MOC) system (Guinan, 2010, 2018; Jennings and Strickland, 2012; Yasin et al., 2014). Apparently consistent with the hypothesis, the presence of a precursor in Tabuchi et al. (2016) removed the MR, which was interpreted as a support for an important role of compression.

More recent evidence, however, argues against a role of efferent compression control in the MR. An otoacoustic emission (OAE) study using 100-Hz amplitude-modulated broadband noise suggested that the MOC activation does not depend on the modulation depth (Mishra and Biswal, 2019), which is consistent with the lack of MOC elicitation observed under various envelope modulation conditions using SPHC stimuli (Wojtczak et al., 2015). Moreover, the finding of MR with both forward and backward maskers (Carlyon et al., 2017) tempers the interpretation of efferent compression control in the MR. The precursor effect in Tabuchi et al. (2016) is, therefore, probably not attributable to the reduction of cochlear gain by MOC activation but, rather, to some other effects such as the dependency on long-term regularity of the temporal envelope (e.g., Münkner et al., 1996; Hickok et al., 2015; Tabuchi and Laback, 2020). Finally, Tabuchi et al. (2016) observed in a preliminary experiment (page 2683) that the MR for SPHC stimuli increases with increasing duration of both masker and target. In the light of the above-described findings on the MOC activation with this type of stimuli, this duration effect appears unlikely to be explainable by an impact of efferent activation, making it reasonable to hypothesize that increasing the stimulus duration increases the MR by means of temporal integration of dip listening.

Thus, in the current study we addressed the duration effect of masker and target more thoroughly, attempting to explore the efficiency of temporal integration in dip listening. Here, dip listening for maskers with more than one dip raises the question how efficient the auditory system combines information across multiple dips for modulated maskers compared to unmodulated maskers. According to the idea of dip-listening integration, a greater MR for modulated compared to unmodulated maskers is expected as the stimulus duration increases. While some temporal integration of a target may occur even with an unmodulated noise masker (Oxenham et al., 1997; Gleich et al., 2007), studies suggest that target integration is more efficient for a modulated than for an unmodulated noise masker (Schooneveldt and Moore, 1989; Gleich et al., 2007). One temporal mechanism for this could be "multiple-looks" integration (Viemeister and Wakefield, 1991), which proposes that the auditory system's sensitivity (d-prime) is determined by integration of information across multiple independent time windows imposed on particular stimulus features. Thus, one may assume that integration across multiple epochs of high target-to-masker energy ratio, i.e., masker dips, is an efficient means for target detection. Such an integration mechanism may be less efficient for a masker with a shallow envelope featuring less pronounced dips, probably because there is more uncertainty of the optimal time epochs to detect the target. Note, however, that simultaneous masking by harmonic maskers with phase relations producing different amounts of envelope fluctuation was reasonably predicted by an excitation-based model involving basilar membrane filtering and compression, which was interpreted as suggestion that temporally selective listening in the dips is not needed to explain those data (Alcántara et al., 2003).

To gain further insights into the impact of dip listening, temporal integration in the MR for simultaneous masking was studied here using deterministic masker stimuli, i.e., SPHCs (Schroeder, 1970), with phase relations producing graded amounts of effective envelope fluctuation (i.e., amplitude modulation depth). In order to facilitate our understanding of the physiological mechanism underlying these temporal effects in the MR, a modeling analysis was carried out based on an established auditory periphery model combined with a detection metric that had been shown to well explain simultaneous masking by reproducible noise, a situation involving temporal detection cues (Richards, 1992; Mao and Carney, 2015). We expected that such a model has the potential to predict target thresholds for maskers with different strengths of envelope fluctuation.

2. Experiment: Effect of covarying masker and target duration

This experiment studied the effect of covarying masker and target duration on the simultaneous MR. By comparing conditions of a short masker and target with conditions of a long masker and target, we intended to test the hypothesis of larger MR for long compared to short stimuli, thus, to replicate the preliminary results of Tabuchi et al. (2016). The short-duration condition was sufficiently short to avoid activation of the efferent system (Backus and Guinan, 2006), and the reduction of compression was thought to be minimal. If, for the long-duration condition, fast-acting compression would be reduced by activation of the efferent system, this would reduce the MR and, therefore, work against the effect of temporal integration of dip listening. Thus, we assumed that any observed increase of the MR with increasing stimulus duration can be attributed only to temporal integration of dip listening.

2.1. Listeners and equipment

Six listeners aged between 18 and 40 years participated in the experiment. All had absolute hearing thresholds of 20-dB hearing level or lower at octave frequencies between 0.25 and 8 kHz. All of the listeners had experience from previous psychophysical masking experiments and received monetary compensation for their participation. None of the authors participated in the experiment. The stimuli were generated on a computer and output via a sound interface (E-Mu 0404, Creative Professional) at a sampling rate of 48 kHz and a resolution of 24 bits. The analog signal was sent through a headphone amplifier (HB6, Tucker-Davis Technologies) to circumaural headphones (HDA 200, Sennheiser). The stimuli were calibrated using an artificial ear (4153, Bruel & Kjær) and a sound level meter (2260, Bruel & Kjær). The experiment was performed in a double-walled sound booth.

2.2. Stimuli

The maskers were Schroeder-phase harmonic complexes (SPHC; Schroeder, 1970; Lentz and Leek, 2001) defined as:

$$m(t) = \sum_{n=N_1}^{N_2} \cos\left[2\pi n f_0 t + \frac{C\pi n(n+1)}{N_2 - N_1 + 1}\right] \quad (1)$$

with $f_0 = 100$ Hz, $N_1 = 16$ and $N_2 = 64$, yielding a masker with a spectral range from 1600 to 6400 Hz. Such a wide masker bandwidth has been typically used in the literature (Oxenham and Dau, 2001b; Shen and Lentz, 2009). The parameter C determines the stimulus phase curvature and controls the peak factor of the acoustic waveform envelope, being maximally peaky for $C = 0$ and flat

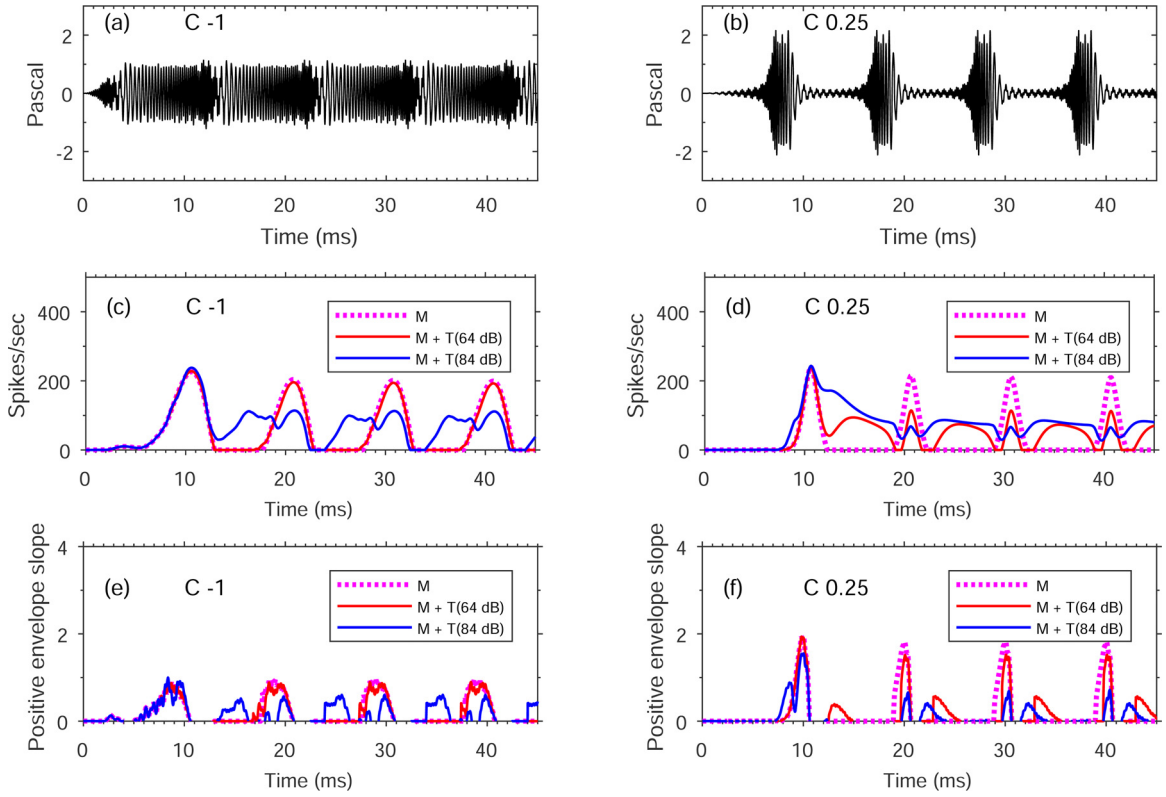

Fig. 1. (Color online) (a) Excerpt of masker waveform with $C = -1$ at 90 dB SPL; (b) Likewise, masker waveform with $C = 0.25$. The "silent dip" of the masker envelope is much longer and lower in level for $C = 0.25$ than $C = -1$ (Mehrgardt and Schroeder, 1983); (c) Example of model's CN spike pattern for masker only (M) and masker plus target ($M+T$) when the masker's C is -1 and the target level corresponds to 64 or 84 dB SPL; (d) Likewise, model's CN spike pattern when the masker's C is 0.25; (e) The positive envelope slope obtained by taking the positive value of the first derivative of the firing pattern of $C = -1$ in panel (c); (f) Likewise, the positive envelope slope obtained from the firing pattern of $C = 0.25$ in panel (d);.

towards $C = \pm 1$. Fig. 1(a) shows the example of a largely flat waveform envelope with $C = -1$, whereas Fig. 1(b) shows a largely fluctuating envelope with $C = 0.25$. With similar stimulus configurations, the highest and lowest masked thresholds have been found with $C = -1$ and $C = 0.25$, respectively (Oxenham and Dau, 2001b; Shen and Lentz, 2009). In this paper, the terms "modulated" and "fluctuating" masker are interchangeably used, both of which describe the presence of pronounced modulation in the temporal envelope, as compared to the "unmodulated" (flat) maskers. Notably, the amount of modulation is controlled by changing the phase spectrum of the SPHC's components without applying any modulator signal. In the experiment, the following six Cs were tested: -1, 0, 0.25, 0.5, 0.75, and 1, based on previous knowledge that the masking effect is maximal for $C = -1$ and minimal for some Cs $\geq$ 0. The target was a pure tone with a frequency of 4000 Hz, spectrally and temporally centered at the masker. The target was added in phase to the masker component at 4000 Hz. Two stimulus durations were tested, referred here to as "Short" and "Long". The Short condition had a masker duration of 40 ms and a target duration of 30 ms (as in Tabuchi et al., 2016), whereas the Long condition had a masker duration of 320 ms and a target duration of 310 ms (as in Oxenham and Dau, 2001b; Shen and Lentz, 2009). The masker-onset to target-onset interval and the target-offset to masker-offset interval were therefore 5 ms. The target and masker were gated on and off with 5-ms cosine-squared ramps. The masker was presented at an overall sound pressure level of 90-dB SPL (*re* 20 µPa). Oxenham and Dau (2001b) and Shen and Lentz (2009) found the MR to be largest for such a high masker level.

Fig. 2 schematically illustrates the temporal characteristics of the stimuli for the two conditions: (a) Short, (b) Long. Additionally,

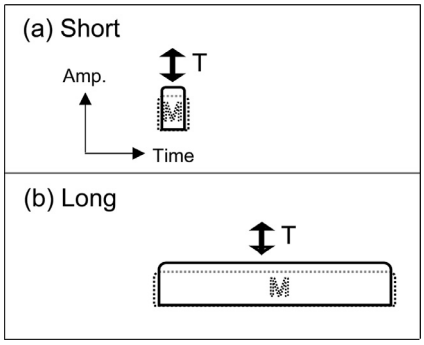

Fig. 2. Schematic of temporal stimulus characteristics used in the experiment. (a) Condition Short; (b) Condition Long. The durations of target and masker were 30 and 40 ms in condition Short, respectively, and 310 and 320 ms in condition Long, respectively. The masker level was 90 dB. The double-headed arrows illustrate the target's amplitude varying during the adaptive threshold measurement. See text for details about the stimuli.

the short and long target durations were tested without a masker. The conditions for the target's absolute threshold (referred to as Quiet) were tested in order to compare the temporal integration in masking to the temporal integration in quiet.

Continuous background noise was added to mask low-frequency cochlear distortion products. The background noise was generated by low-pass filtering a Gaussian white noise with a second-order Butterworth filter (12-dB/oct attenuation, cut-off frequency of 1300 Hz) and presented at an overall SPL of 70 dB. Stimuli were presented to the listeners' right ear.

2.3. Experimental procedure

An adaptive three-interval forced choice procedure with a three-down one-up staircase rule was used to measure masked and absolute target thresholds at 79% correct (Levitt, 1971). The silent intervals between the three stimuli of a trial were 700 ms. This duration was considered as sufficient to account for the time constant of MOC activation decay (Backus and Guinan, 2006; Walsh et al., 2010). The masker was presented in all three intervals, whereas the target was presented only in one randomly selected interval. The listeners indicated the interval which sounded different from the other two by pressing the corresponding button. Feedback on the correctness of the response was provided visually after each trial. Each run was terminated after 8 reversals. At the beginning of each threshold run, the target was presented at a sound level clearly above the expected threshold. The step size of adaptive level change was 4 dB up to the fourth reversal, and then reduced to 2 dB up to the eighth reversal. The target threshold was estimated from the average of the last four reversals.

The order of testing was blocked according to the duration condition (Short or Long). Within each block, threshold measurement for each C was conducted three times, randomizing the order of Cs and repetitions. Listeners moved on to the other duration condition after one duration condition was completed. The order of duration conditions were random for each listener. All Quiet conditions were run after the SPHC masking conditions were completed. Listeners were allowed to take a break every three threshold measurements, if necessary. The total testing time of the experiment amounted to 3 to 4 h per listener.

2.4. Modeling analysis procedure

To obtain further insight into the mechanism underlying temporal integration of MR, we conducted a physiology-based modeling analysis. We applied a variant of the so-called envelope-slope metric (ESM; Richards, 1992; Mao and Carney, 2014, 2015), a measure of the stimulus-evoked fluctuation strength, to the stimulus representation as determined by a model of the auditory periphery up to the level of the cochlear nucleus (CN). The model of the auditory periphery up to the auditory nerve (AN) level is well established (Carney, 1993; Zilany et al., 2009, 2014), accounting for many properties of AN responses, including short-term and long-term adaptation, decay of forward masking, and some basic features of the temporal modulation transfer function (TMTF). Recent models of the CN and the inferior colliculus (IC), added to the AN model, further elaborated the characteristics of the TMTF based on excitatory and inhibitory responses (Nelson and Carney, 2004; Carney et al., 2015; Carney and McDonough, 2019). We applied the ESM to the output of the CN because the CN is the first model stage removing the temporal fine structure of AN firing patterns and thus extracting the temporal envelope.

Following Mao and Carney (2015), we implemented the ESM as the integrated positive gradient of the envelope signal, i.e., in our case the CN response. Figs. 1(c) and 1(d) show, as examples, the CN firing patterns of a weakly modulated masker M ($C = -1$) and a highly modulated masker ($C = 0.25$) both without a target (dotted line) and with a target T at levels of 64 dB and 84 dB SPL (solid line), covering the entire level range of masked threshold measurements. Remarkably, periodic temporal dips occur in the firing pattern for $C = -1$ and the dips start being filled by the target only at the higher level. In contrast, the dips for $C = 0.25$ are already filled at the lower target level.

Panels (e) and (f) of Fig. 1 show the positive envelope slope across time, obtained by half-wave rectification of the first derivative of the firing pattern for $C = -1$ and $C = 0.25$ in panels (c) and (d), respectively. The positive envelope slope generally provides a measure of "the envelope fluctuation strength" (e.g., Klein-Hennig et al., 2011; Mao and Carney, 2015). When the target tone is added, the fluctuation strength decreases particularly around the envelope peaks of the masker (e.g., around ~ 21, 31, and 41 ms) for both $C = -1$ and $C = 0.25$, though the decrease is more pronounced for the high target level than for the low target level (T 84 dB vs T 64 dB), and is more pronounced for the condition with more envelope fluctuation than for the condition with less envelope fluctuation ($C = 0.25$ vs $C = -1$). This effect appears to result from the decline in the F0-based response relative to the target-frequency response, referred to as synchrony capture (see, e.g. Carney et al., 2015; Carney, 2018; Carney and McDonough, 2019). In contrast, in the envelope dips of the masker (e.g., ~ 13, 23, and 33 ms), increasing the target level, slightly and gradually *increases* the fluctuation strength. The net effect of strong decrease and slight increase across the masker period is an overall *decrease* of fluctuation strength with increasing target level [see Figs. 3(a) and 3(b)]. Our decision variable, the "ratio of integrated slopes" (RIS), first integrates the positive slope for $M+T$ [i.e., notated as $\Sigma\mathrm{ESM}(M+T)$] and for M [i.e., $\Sigma\mathrm{ESM}(\mathrm{M})$] over the stimulus duration, and then determines the ratio $\Sigma\mathrm{ESM}(M+T)/\Sigma\mathrm{ESM}(\mathrm{M})$. The idea behind the RIS metric is that the listener builds a representation of M alone when listening to the intervals without a target and compares this to the representation of $M+T$ when listening to the target interval. Figs. 3(a) and 3(b) plot RIS as a function of the input target level for $C = -1$ and $C = 0.25$ in the Short and Long condition, respectively, thus generating a decision-variable function (DVF) for each condition. The shape of the DVF is critical for the predicted masked threshold, as determined by the RIS criterion (horizontal dotted line). In the example conditions of Figs. 3(a) and 3(b), the DVF decreases faster with increasing target level for $C = 0.25$ than for $C = -1$, resulting in a lower predicted threshold for $C = 0.25$. In other words, adding the target to the masker results in a reduction of fluctuation strength, and the reduction is stronger for $C = 0.25$ than for $C = -1$. The reduction of RIS with increasing target level is most pronounced for $C = 0.25$ in the Long condition, compared with the other conditions, suggesting that the RIS metric is sensitive to the integration of reduced envelope fluctuation. In order to predict masked thresholds for a given set of data, the criterion RIS was systematically varied as the only free model parameter, minimizing the root-mean-square error (RMSE) between mean experimental thresholds and corresponding predictions. All the RMSEs and criterion RISs estimated in the present study are listed in Table II.

Throughout this paper we report predictions obtained with Low-SR (spontaneous rate) AN fibers. Over the course of our modeling analysis, Low-SR fibers seem to show better prediction accuracy (in terms of RMSE) than the Medium-SR or High-SR fibers.[2] Each point of the DVF represents the average RIS across 10 model repetitions. The background noise used for the experiment was not added to the model's input stimuli. The parameter of characteristic frequency (CF) was fixed at 4000 Hz at the AN modeling stage, and it was the sole CF parameter tested in the current study. We used the default model parameters of AN and CN as given by the latest model version (UR EAR 2.0) and listed in the Appendix. Those CN parameters were chosen to avoid excessive temporal synchronization to the stimulus onset and offset

[2] At the 90-dB masker level used for our experiments, the firing rates of High-SR and Med-SR fibers are saturated and therefore not sensitive to stimulus manipulations. It is a common physiological practice to focus on Low-SR fibers for high levels. High-SR and Med-SR fibers may actually be sensitive to target information in the masker's dip, but we wanted to avoid additional assumptions on physiological mechanisms within the scope of this study. A recent perspective proposes that High-SR fibers associated with IHC-AN synapses may benefit coding spectro-temporal fluctuation profiles (see Carney, 2018).

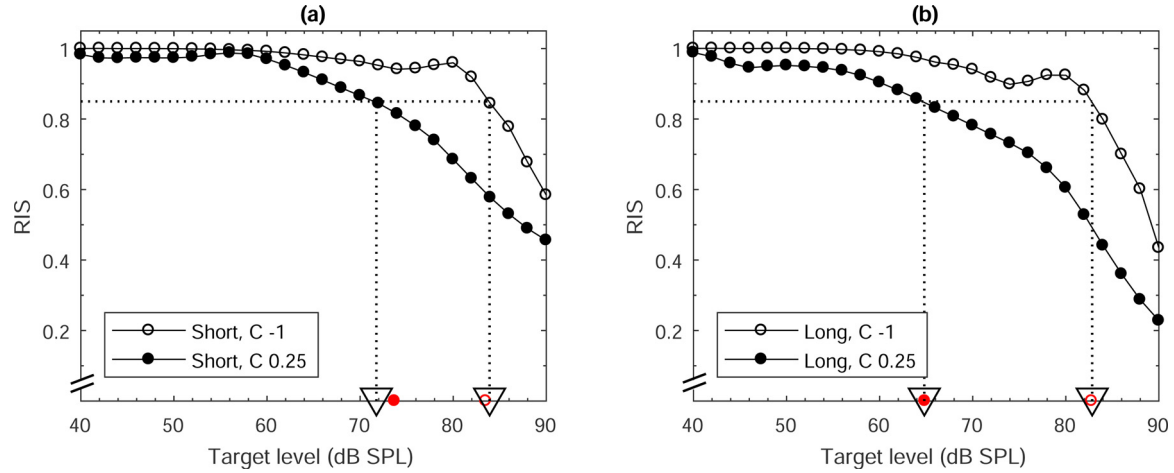

Fig. 3. (Color online) Ratio of integrated positive envelope slope (RIS) as a function of target level. (a) Condition Short; (b) Condition Long. To compute RIS, the positive envelope slope was first integrated across the entire stimulus duration for both $M+T$ and M, and then the integrated positive envelope slope of $M+T$ was divided by the integrated positive envelope slope of M. This computation was repeated across target levels for each condition, resulting in a decision-variable function (DVF) shown in each panel. The points in 2-dB steps were linearly interpolated, and the inverse of a criterion RIS (here: 0.85) was defined as the predicted threshold (vertical arrows). The mean masked threshold in each condition is shown by the circles on the x-axis. The criterion RIS was found by minimizing the root-mean-square error (RMSE) between mean experimental thresholds and predicted thresholds (see Table II). See text for details about the simulation.

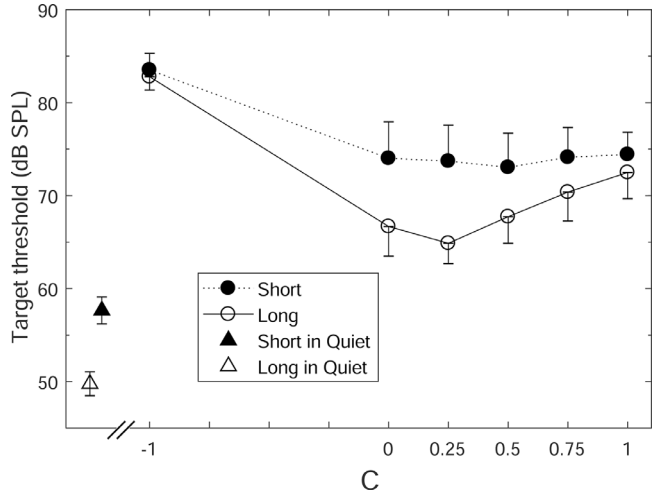

Fig. 4. Mean target thresholds as a function of *C* for conditions Short and Long and for those in Quiet (without masker). Note, that the background noise was present even in the "Quiet" condition. Error bars indicate ±1 standard error across the six listeners.

Table I
Listener-specific masking release (MR, i.e., differences in dB between the target thresholds for *C*s −1 and 0.25) for different stimulus durations. The bottom row shows the MR obtained from the averaged thresholds across the six listeners.

Listener	Short	Long
NH14	17.7	19.5
NH39	7.3	13.8
NH43	12.5	19.8
NH143	3.0	18.5
NH144	4.0	15.5
NH714	14.3	20.3
Mean	9.8	17.9

(Mao and Carney, 2015), which was also desirable for our purpose. Further optimizing the CN parameters would have exceeded the scope of the present study.

2.5. Results and discussion

Fig. 4 shows mean masked thresholds across listeners as a function of *C* for conditions Short and Long, respectively. Starting with condition Long (open symbols), the masked threshold is highest for $C=-1$ and lowest for $C=0.25$. We quantified the difference of thresholds between those extreme values as a measure of MR, supported by extreme values found at similar *C*s in published studies (Oxenham and Dau, 2001b; Shen and Lentz, 2009). For condition Long, the mean MR across listeners amounts to 17.9 dB (see also Table I for listener-specific MRs). For condition Short (filled symbols), the masked threshold is also the highest for $C=-1$ and very similar to condition Long, but for *C*s $\geq$ 0 the thresholds are approximately constant (around 74 dB SPL) and systematically higher than in condition Long (resulting in a mean MR of 9.8 dB). Note that the 8.1 dB difference in MR between conditions Short and Long is almost entirely due to the decreased masked thresholds for the long duration at the stimulus phase curvatures where the stimulus envelope is highly fluctuating. Overall, these effects were supported by the results of a two-way repeated-measures analysis of variance (RM-ANOVA), showing significant main effects of both the factors *C* [$F(5,25)=22.56$, $p<0.001$] and stimulus duration [$F(1,5)=7.684$, $p=0.039$] and a significant interaction between them [$F(5,25)=7.773$, $p<0.001$].

When the modeling analysis was conducted across all the *C* values, the model predictions were found to largely deviate from the mean experimental thresholds (see Table II for the RMSE). Despite the failure of predictions based on a single RIS criterion across the six *C*s, the model coincidentally predicted the overall patterns of maximum vs. minimum thresholds [for *C*s of −1 and 0.25 as shown in Figs. 3(a) and 3(b), respectively] for conditions Short and Long. Because the current study is primarily concerned with the duration effect on the MR difference between internally weakly versus highly modulated stimuli (i.e., *C*s of −1 and 0.25), and assuming that the model adequately represents that "relative" duration effect, our subsequent modeling focuses on the prediction of such defined MR rather than the prediction across all the six *C*s.

Fig. 5(a) shows the mean masked thresholds and model predictions optimized for the two selected *C*s of −1 (open symbols) and 0.25 (filled symbols) and the two duration conditions (measured thresholds replicated from Fig. 4). Fig. 5(b) shows the corresponding MRs. The model well captures the pattern of the duration effect (see also, Table II). To discern the potential influence of onset and

Table II
Estimates of best fitting "ratio of integrated slopes" (RIS) criterion and corresponding root-mean-square errors (RMSE) in dB between the mean thresholds and model predictions. (Upper row) Criterion and RMSE obtained when predicting *all* conditions, i.e., *six Cs* and the two stimulus durations; (Middle) Criterion and RMSE obtained when predicting *two Cs* and the two durations; (Bottom) Criterion and RMSE obtained when predicting *two Cs* and the two durations when the onset and offset of CN responses (spikes/*sec*) were excluded from the computation (denoted as "Trimmed"). See text for details about the model.

C	Duration	Fig.	RMSE	RIS criterion	Remarks
−1, 0, 0.25, 0.5, 0.75, 1	Short, Long	N/A	9.69	0.85	DVF for C −1 & 0.25 in Fig. 3
−1, 0.25	Short, Long	5	0.80	0.84	
−1, 0.25	Short, Long	5	1.87	0.81	Trimmed

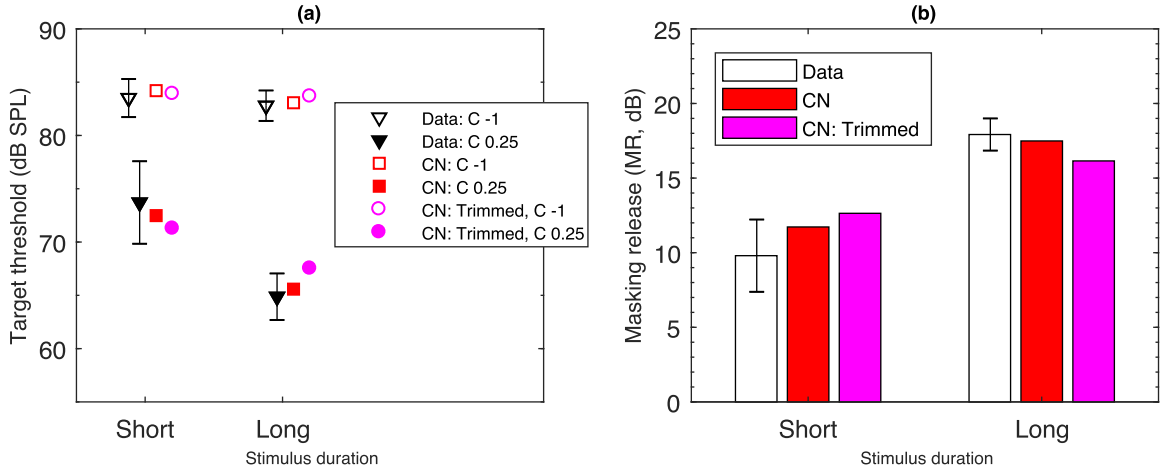

Fig. 5. (Color online) (a) Mean target thresholds for the stimulus durations (Short and Long) and for the *Cs* (−1 and 0.25), and corresponding CN-model predictions. The predictions excluding onset and offset portions are denoted as "Trimmed". Predictions were performed by optimizing the RIS criterion for the given set of data (see Table II); (b) Mean masking release (MR; i.e., the threshold difference between the *Cs* of −1 and 0.25) corresponding to conditions from panel (a). Error bars indicate ±1 standard error across the six listeners. The listener-specific MRs are shown in Table I. See text for details about the simulation.

offset portions on the model prediction, we tested the predictions based on the "trimmed" CN responses as shown by the circles. The trimming operation removed the CN response corresponding to the first and last masker periods, and used only the CN response between 20 and 30 ms after stimulus onset for condition Short (i.e., 1 cycle) and between 20 and 310 ms for condition Long (29 cycles). The predictions for trimmed responses (circles) differ only slightly from those for the complete responses without trimming (squares), suggesting no significant contribution of the onset and offset to the prediction.[3]

To illustrate the dependency of the ESM (leading to the RIS) on stimulus duration, Fig. 6 shows the sum of the ESM across stimulus duration for target levels of 64 dB [panel (a)] and 84 dB SPL [panel (b)], respectively. The masker-only (M) conditions for $C=-1$ and $C=0.25$ (triangles) are the same in both panels, serving as baselines. The first important observation is that the sum of ESM for $M+T$ (filled or open circle) is lower for the high target level [Fig. 6(b)] than for the low target level [Fig. 6(a)]. Second, the sum of ESM for $M+T$ (the circles) increases less with increasing stimulus duration, compared to the sum of ESM for M (the triangles). Third, the reduction of sum of ESM for $M+T$ relative to the sum of ESM for M is stronger for the highly modulated ($C=0.25$) masker than for the weakly modulated ($C=-1$) masker. To summarize, the sum of ESM is lowest when the target level is high (84 dB) [filled circle in Fig. 6(b)], when the stimulus duration is long (320 ms), and when the masker is modulated ($C=0.25$). These fine-grained computations of the sum of ESM are consistent with the general description in Section 2.4. Comparison to Fig. 1 suggests that the variation of sum of ESM with target level and stimulus duration is dominated by the changes in firing pattern [Fig. 1(c) and Fig. 1(d)] and positive slope [Fig. 1(e) and Fig. 1(f)] occurring particularly during the temporal *peak* of the masker.

Finally, Fig. 7 compares the mean size of experimentally observed temporal integration in conditions of masking ($C=-1$ vs $C=0.25$) and in quiet (without masker). The size of temporal integration is very similar for the modulated masker ($C=0.25$) and in quiet, while it is rather small for the unmodulated masker ($C=-1$). An RM-ANOVA performed on the size of temporal integration indicated significant main effects of the factor masker configuration [$F(2,10)=13.6$, $p=0.001$]. Post hoc pairwise comparisons using the Tukey LSD test indicated a significant difference between $C=-1$ and $C=0.25$ ($p<0.01$) and between $C=-1$ and Quiet ($p<0.001$) and, more importantly, no significant difference between $C=0.25$ and Quiet ($p=0.674$). This suggests that the size of temporal integration of a target tone masked by a highly modulated SPHC masker is not significantly different from that of a target tone in quiet. While this comparison demonstrates a comparable size of temporal integration, it does not mean that the underlying integration mechanism is the same. For a highly modulated masker ($C=0.25$), listeners might follow a "multiple-looks" approach, combining target information from individual masker dips (i.e., "looks"). Note that, in the ESM model, each look corresponds to a change in the integrated positive envelope slope within one masker period. For an unmodulated a single pure tone in quiet, listeners may instead simply integrate energy across the entire stimulus duration up to ~300 ms according to an "energy detector" (e.g., Plomp and Bouman, 1959; Green, 1960). More generally, in

[3] The temporal decay of CN firing-rate across stimulus duration (i.e., short adaptation) seems similar between the masker fluctuation conditions ($C=-1$ vs. 0.25) within each duration condition, and the difference of temporal decay between the masker fluctuation conditions does not appear to change with the duration conditions (Short vs Long). In our modeling analysis, it is thus unlikely that short-term adaptation contributed to the stimulus duration effect.

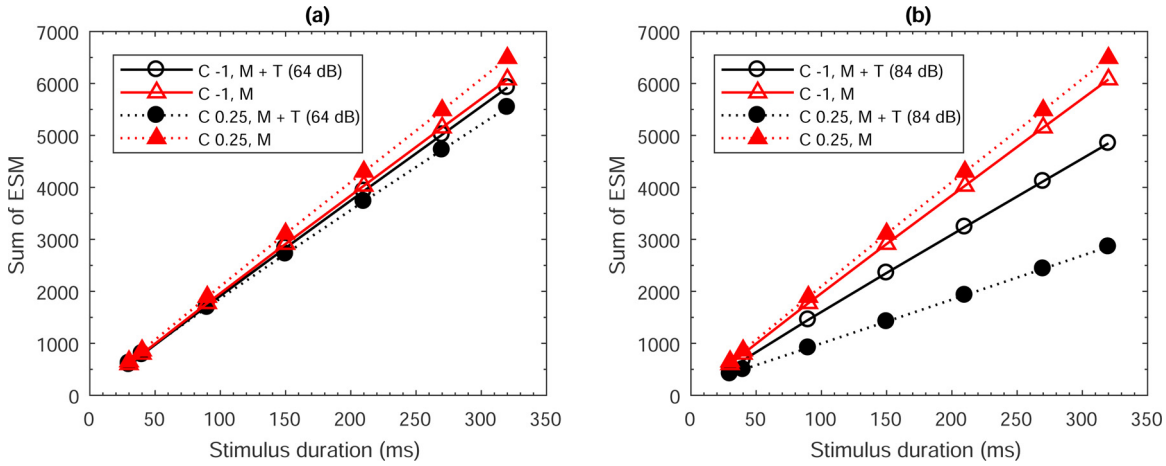

Fig. 6. (Color online) Sum of the model's envelope-slope metric (ESM) as a function of stimulus duration. Each panel shows the sum of ESM for target levels of 64 dB SPL [panel (a)] and 84 dB SPL [panel (b)]. ESM decreases when adding the target (T) with an increasing level [panel (a) vs. (b)], and when increasing the masker's (M) envelope modulation ($C = 0.25$ vs. -1). The triangles for masker alone (M) serve as baselines in both panels. The stimulus duration on the horizontal axis denotes the masker duration (i.e., 30, 40, 90, 150, 210, 270, and 320 ms), with the target duration being 10 ms shorter than the masker duration. All stimulus variables (e.g., masker level 90 dB SPL) are the same as those used in the experiment. See text for details about the simulation and interpretation of the figure.

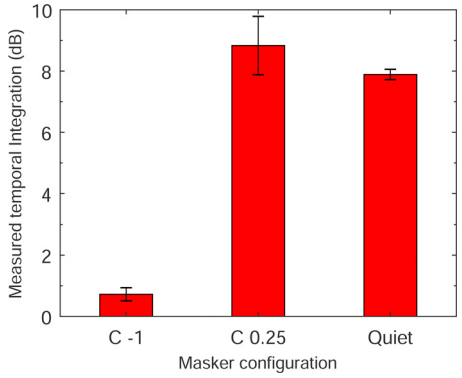

Fig. 7. (Color online) Mean size of temporal integration for conditions with a masker (*C*s of -1 and 0.25), and in quiet (without a masker), quantified as the threshold difference between conditions Short and Long in dB. Error bars indicate ± 1 standard error across the six listeners.

masking situations listeners may "flexibly" follow different strategies of integration, depending on the temporal structure of the masker (see Plack, 2018). According to our model assumption, the listening strategy might be similar for a weakly modulated masker as for a highly modulated masker. In fact, for the Short condition, the change in the integrated positive envelope slope in each masker period for the highly modulated masker ($C = 0.25$) is only slightly smaller than for the weakly modulated masker ($C = -1$), resulting in essentially very similar thresholds. However, for the Long condition, the small difference accumulates across time such that the sum of slope across all periods becomes considerable larger for the highly modulated masker than for the weakly modulated masker (see Fig. 6), resulting in a lower threshold compared to the weakly modulated masker.

3. General discussion and conclusions

It is well established that the amount of simultaneous masking by a Schroeder-phase harmonic complex (SPHC) depends strongly on its effective envelope fluctuation. A masker with a strong envelope fluctuation ($C = 0.25$) produces less masking than a masker with a weak envelope fluctuation ($C = -1$), leading to masking release (MR). The present study attempted to shed light on the potential mechanisms contributing to the MR obtained with this type of stimulus by measuring the effect of overall stimulus duration and predicting the results using a model that combines a state-of-the-art auditory periphery frontend with a temporal-envelope based decision metric. In contrast to previous studies on temporal integration in MR using continuous maskers (Schooneveldt and Moore, 1989; Gleich et al., 2007), we used gated maskers to obtain information about the potential contribution of activation of the efferent system (see below). Moreover, in contrast to studies on target's temporal integration using gated maskers with a fixed duration (e.g., Oxenham et al., 1997), we covaried the duration of both masker and target to avoid the potentially confounding influence of the masker's trailing portions (before target onset and after target offset), which allowed us to keep constant any effects resulting from those target-free portions of the masker.

Our study showed a significantly larger MR for the long compared to the short stimulus condition, i.e., demonstrating temporal integration in the MR. The results confirmed the preliminary findings of Tabuchi et al. (2016) and provide further evidence for stronger temporal integration of target information for a highly modulated masker compared to a weakly modulated masker. To our knowledge, evidence for this has so far been provided only for continuous noise maskers, either as a byproduct of a study on co-modulation masking release (Schooneveldt and Moore, 1989) or in a behavioral animal study with gerbils (Gleich et al., 2007).

Because fast cochlear compression is thought to contribute to the simultaneous MR (e.g., Oxenham and Dau, 2001a, 2001b, 2004; Alcántara et al., 2003), it was considered that a reduction of this compression over the course of the stimulus by means of activation of the efferent system might reduce the contribution of later stimulus portions to the MR (see Tabuchi et al., 2016). This should cause either no effect or even a smaller MR for long compared to short stimuli. However, the finding that the MR in fact *increased* with increasing stimulus duration suggests that an efferent-induced reduction of compression either had no effect or only a small effect which was overruled by another mechanism, namely temporal integration of dip listening. Notably, this does not mean that fast compression *per se* does not contribute to the MR. Interestingly, an increase of MR with increasing stimulus duration had been observed also with a forward masking paradigm (Wojtczak and Oxenham, 2009), leading to the conclusion that efferent compression control is unlikely to play an important role in the masker phase effect (Wojtczak et al., 2015). For a further discussion of results obtained in simultaneous versus forward masking, see below.

Despite these arguments, the current study does not generally rule out the involvement of MOC activation in the MR. In OAE studies, there has been a long debate whether various envelope fluctuations elicit MOC activation or not (e.g., Guinan, 2010); for example, MOC activation varies with stimulus variables of SPHCs, such as the waveform crest factor, the F0, and the envelope rate (Micheyl et al., 1999). If the MOC system was activated in the current study for tangled or unknown reasons, it is not unimaginable that time-evolving MOC activation could have reduced the frequency selectivity and thus expanded the AF bandwidth (e.g., Strickland, 2001) towards the end of long $M+T$, which may allow more frequency components to fall within the the auditory filter (AF) and increase the envelope fluctuation, finally resulting in a large MR. On the other hand, at the high masker level used in the present study, there would not be much room for a widening of AF bandwidth.

Together, the current simultaneous masking results appear consistent with the idea that listeners efficiently integrated target information across masker dips for the highly modulated masker, whereas the lack of such informative epochs prevented such a strategy for the less modulated masker (see below). As a control condition, temporal integration was also measured without masker, using the same target as in the masked conditions. Interestingly, the amount of temporal integration in quiet was found to be equal in magnitude to temporal integration in presence of the modulated masker ($C=0.25$). This indicates that listeners can be comparably efficient in integrating short bits of information across time as they are in integrating continuously available information.

The results of this study can be partly interpreted in terms of the multiple-looks model of temporal integration (Viemeister an Wakefield, 1991; Donaldson et al., 1997), assuming that each masker dip, i.e., temporal epoch of high target-to-masker energy ratio, represents a "look". An important rationale of that model is that the amount of threshold improvement associated with an increase in the number of looks affects the slope of the psychometric function. Steep psychometric functions will result in less threshold improvement than shallow psychometric functions (Viemeister an Wakefield, 1991), presumably because a difference of target tone levels corresponding to a constant improvement in performance is smaller in case of a steeper function. However, to better interpret the current data in terms of the multiple-looks integration, a rigorous measurement of the psychometric function's slope is required, thus requiring the use of the method of constant stimuli in a future study.

In order to gain insight into the mechanism involved in the MR, we predicted the masked thresholds with a physiology-inspired model of the auditory processing up to the CN. Based on the established envelope slope metric (ESM, Richards, 1992; Strickland and Viemeister, 1996; Mao and Carney, 2015), the modeling approach assumed that in each trial listeners evaluate the difference in the "internal" (neural) envelope patterns between the masker alone (M) and the masker plus target ($M+T$). The model idea is that listeners detect a reduction of overall envelope fluctuation strength when adding the target to a modulated masker. Specifically, our version of the metric, the RIS, compared the summed positive instantaneous slope of the CN's firing pattern for M alone with that for $M+T$. The RIS metric was able to predict the duration effect in the MR, as determined based on the masker causing maximum masking ($C=-1$) versus the masker causing minimum masking ($C=0.25$) in short versus long stimulus durations. At a more detailed level, it was found that in order to obtain these predictions, a critical property of the CN's model response to $M+T$ was a decline of the positive envelope slope during the response peak when adding a target with increasing level [see Figs. 1(e) and 1(f)]. While there is also a counteracting increase of the positive envelope slope during the response dip, the decline during the peak was found to dominate the increase during the dip. The decline of the positive envelope slope during the response peak was ascribed to the well-known effect of synchrony capture (see, e.g. Carney et al., 2015; Maxwell et al., 2020). In other words, dip listening, and therefore also temporal integration across multiple dips, for SPHC stimuli appears to rely on the reduction of the peak in the neural envelope representation within each masker period. While this might appear contradictory, this just means that at the limits of perception, i.e., at masked threshold, the auditory system seems to rely on features of the neural signal representation that are only indirectly related with "external" (acoustic) signal, due to the mechano-electrical transformations in the auditory periphery. It should be mentioned, though, that such a strategy only works for a simple detection task, but it would not work for signal recognition (e.g. speech) in temporal dips.

On the one hand, the model did predict the minimum and maximum masked threshold (across *C*s) for conditions Short and Long, which were found to be consistent with measurements from the literature for such stimuli and were selected to determine the MR before starting with the model analysis. On the other hand, for other C values beyond this preselected data set ($C=-1$ and $C=0.25$), the model did not predict the pattern of masked thresholds using a constant threshold criterion, suggesting that the good absolute predictions for threshold maxima and minima were coincidental. Interestingly, inspection of the DVF for all other *C*s (not shown) suggested that the area covered between short- and long-stimulus DVFs is proportional to the amount of the duration effect for each *C*. In other words, an integrated quantity of RIS rather than a single RIS criterion appears to be well correlated with the duration effect even across various *C*s, possibly because such integrated quantity is less sensitive to the absolute dB SPL of masked thresholds. Overall, however, the current model demonstrates the lack of predictability of the detailed threshold pattern across *C*s which appears to be due to a discrepancy in the phase response of the AN frontend model and that of actual listeners (see also Tabuchi and Laback, 2017). Thus, there remains room for adapting the AN model's phase response to better reflect masker phase effects for SPHC stimuli. Moreover, low-SR fibers were found to be by far most predictive of masking effects, leaving open the question how the auditory system combines information from the three AN fiber types (Low-, Med-, and High-SR) to detect a tone masked by temporally fluctuating maskers.

We also applied another decision metric, namely the integrated d-prime (Pressnitzer et al., 2001) which is conceptually equivalent to the decision metric (d-prime) used for analyzing multiple-looks integration in human psychoacoustics (Viemeister and Wakefield, 1991). One problem we observed when applying this metric to the model's CN firing pattern was that the firing rate for the highly modulated masker ($C=0.25$) can reach zero at the masker's dip, especially when the analysis window is short ($< \sim 4$ ms), such that the predicted sensitivity within the corresponding short time window becomes excessively or infinitely high, by far exceeding human performance and, thus, resulting in inaccurate predictions. While this issue may be resolved by adding an internal noise source, another issue was the violation of "normality" of the probability density function based on the firing rates within short time windows. Namely, the variance of the probability density function was sometimes extremely large for highly modulated maskers ($C=0.25$), due to the impulsive character of the corresponding firing-rate pattern. In some cases, the probability density function was even bimodal (i.e., having two peaks), making it hard to robustly determine d-prime scores. Because of these problems, we aborted our attempts to predict the results based on the d-prime metric.

Compared to the integrated d-prime metric, the RIS metric was found to be computationally more feasible. A particularly attrac-

tive aspect of the RIS metric is that it does not require the specification of particular short time windows, such as in d-prime based models (e.g., Pressnitzer et al., 2001). For example, Mao et al. (2013) showed that an ESM model version that segmented the stimulus into equal-duration windows achieved the same prediction power for tone-in-reproducible-noise detection as a version using a single long-window. Note, however, that the RIS metric is limited to simultaneous configurations, and it is unlikely to explain the masked threshold in forward masking (Wojtczak and Oxenham, 2009) and backward masking (Carlyon et al., 2017).

From a broader perspective, the demonstrated increase of simultaneous MR with increasing stimulus duration appears at first sight consistent with an increase of nonsimultaneous MR with increasing masker duration observed in a forward masking paradigm (Wojtczak and Oxenham, 2009). But, because dip listening is not possible in a forward masking configuration, this apparent similarity of effects may raise concerns about our interpretation of dip listening in simultaneous MR. On the other hand, given the rather different nature of stimulus variations in the simultaneous and forward masking studies (we varied both masker and target duration, while Wojtczak and Oxenham, naturally had to keep the target duration constant), it is not unlikely that different mechanisms are involved in those two masker configurations. Although beyond the scope of the present study, one potential mechanisms for the masker duration effect in forward masking could be that listeners entrained to the rhythm of a modulated masker and the persistence of this rhythm at a more central level causes some kind of higher-order dip-listening effect (Hickok et al., 2015). Increasing the masker duration may improve this form of dip listening due to stronger entrainment, resulting in larger MR. Taken together, while the observed duration dependence of the simultaneous MR and its modeling appears to be consistent with an explanation in terms of dip listening (in the "external" signal) and synchrony capture (in the neural signal), the contribution of other mechanisms, which might also explain the corresponding duration effects in forward masking, cannot be ruled out.

To summarize, the following main conclusions can be drawn from the current experimental and modeling study. The masking release for highly versus weakly fluctuating (modulated) Schroeder-phase harmonic complex stimuli was found to increase with increasing masker and target duration. The duration effect in the masking release was predicted by the envelope slope metric (ESM) applied to an auditory periphery model up to the cochlear nucleus. The duration was interpreted in terms of temporal integration of target information across masker envelope dips (at the signal level) and in terms of temporal integration of a decline in the neural envelope peak (i.e., synchrony capture) across modulation periods (at the neural level). These interpretations of the duration effect are, thus, different manifestations of the same phenomenon.

Future research is required to study the generalization of the current findings to a broader range of stimuli with various envelope modulation characteristics. Using a frontend AN model that better reflects the human cochlear phase response might advance the predictability of target detection in presence of arbitrary maskers with given phase characteristics.

Author statement

Hisaaki Tabuchi: Conceptualization, Data acquisition, Data analysis, Writing. Bernhard Laback: Conceptualization, Writing.

Acknowledgments

We would like to thank the two anonymous reviewers for providing helpful comments on an earlier version of the manuscript. We also thank Professor Armin Kohlrausch for inspiring discussion, and Michael Mihocic for his assistance in writing software programs for the experiments (*ExpSuite*). The first author is grateful to Professor Pierre Sachse for his constant encouragement and patience. This work was supported in part by the Austrian Science Fund (FWF, P24183-N24), by the Tirol Science Fund (TWF, P7200–045–011), and by an award from the University of Innsbruck (Nachwuchsförderung).

Appendix

In the modeling analysis, we used the following parameters:
AN model:

- CF: 4 kHz
- Model sampling rate: 100 kHz
- Outer hair cell (OHC) scaling factor: 1 (normal)
- Inner hair cell (IHC) scaling factor: 1 (normal)
- Species: human, using basilar membrane tuning from Shera et al. (2002)
- Fractional Gaussian noise type: variable
- Power-law implementation: approximate
- Number of stimulus repetition (nrep): 1000

CN model:

- Excitatory input time constant τ_{ex}: 0.5 ms
- Inhibitory input time constant τ_{ihn}: 2 ms
- Inhibitory delay D_{ihn}: 1 ms
- Amplitude of inhibition C_{ihn}: 0.6
- Scalar for model stage output A: 1.5
- Best modulation frequency: 100 Hz

References

Alcántara, J.I., Moore, B.C., Glasberg, B.R., Wilkinson, A.J., Jorasz, U., 2003. Phase effects in masking: within- versus across-channel processes. J. Acoust. Soc. Am. 114, 2158–2166.
Backus, B.C., Guinan, J.J., 2006. Time-course of the human medial olivocochlear reflex. J. Acoust. Soc. Am. 119, 2889–2904.
Bacon, S.P., Lee, J., Peterson, D.N., Rainey, D, 1997. Masking by modulated and unmodulated noise: effects of bandwidth, modulation rate, signal frequency, and masker level. J. Acoust. Soc. Am. 101, 1600–1610.
Bacon, S.P., Lee, J., 1997. The modulated-unmodulated difference: effects of signal frequency and masker modulation depth. J. Acoust. Soc. Am. 101, 3617–3624.
Bacon, S.P., Opie, J.M., Montoya, D.Y., 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. J. Speech Lang. Hear. Res. 41, 549–563.
Buus, S., 1985. Release from masking caused by envelope fluctuations. J. Acoust. Soc. Am. 78, 1958–1965.
Buss, E., Hall, J.W., Grose, J.H., 2012. Effects of masker envelope irregularities on tone detection in narrowband and broadband noise maskers. Hear. Res. 294, 73–81.
Carney, L.H., 1993. A model for the responses of low-frequency auditory-nerve fibers in cat. J. Acoust. Soc. Am. 93, 401–417.
Carney, L.H., 2018. Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. J. Assoc. Res. Otolaryngol. 19, 331–352.
Carney, L.H., McDonough, J.M., 2019. Nonlinear auditory models yield new insights into representations of vowels. Atten. Percept. Psycho. 81, 1034–1046.
Carney, L.H., Li, T., McDonough, J.M., 2015. Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. eNeuro 2 (4).
Carlyon, R.P., Datta, A.J., 1997. Excitation produced by Schroeder-phase complexes: evidence for fast-acting compression in the auditory system. J. Acoust. Soc. Am. 101, 3636–3647.
Carlyon, R.P., Flanagan, S., Deeks, J.M., 2017. A re-examination of the effect of masker phase curvature on non-simultaneous masking. J. Assoc. Res. Otolaryngol. 18, 815–825.
Deroche, M.L., Culling, J.F., Chatterjee, M., 2013. Phase effects in masking by harmonic complexes: speech recognition. Hear. Res. 306, 54–62.
Donaldson, G.S., Viemeister, N.F., Nelson, D.A., 1997. Psychometric functions and temporal integration in electric hearing. J. Acoust. Soc. Am. 101, 3706–3721.
Freyman, R.L., Griffin, A.M., Oxenham, A.J., 2012. Intelligibility of whispered speech in stationary and modulated noise maskers. J. Acoust. Soc. Am. 132, 2514–2523.
Gleich, O., Kittel, M.C., Klump, G.M., Strutz, J., 2007. Temporal integration in the gerbil: the effects of age, hearing loss and temporally unmodulated and modulated speech-like masker noises. Hear. Res. 224, 101–114.
Green, D.M., 1960. Auditory detection of a noise signal. J. Acoust. Soc. Am. 32, 121–131.

Green, T., Rosen, S., 2013. Phase effects on the masking of speech by harmonic complexes: variations with level. J. Acoust. Soc. Am. 134, 2876–2883.
Guinan, J.J. (2010). Physiology of the Medial and Lateral Olivocochlear Systems. In: Ryugo D., Fay R. (eds) Auditory and Vestibular Efferents. Springer Handbook of Auditory Research, vol 38. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-7070-1_3
Guinan, J.J., 2018. Olivocochlear efferents: their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hear. Res. 362, 38–47.
Hickok, G., Farahbod, H., Saberi, K., 2015. The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. Psychol. Sci. 26, 1006–1013.
Jennings, S.G., Strickland, E.A., 2012. Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity. J. Acoust. Soc. Am. 132, 2483–2496.
Klein-Hennig, M., Dietz, M., Hohmann, V., Ewert, S.D., 2011. The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. J. Acoust. Soc. Am. 129, 3856–3872.
Kohlrausch, A.G., Sander, A., 1995. Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. J. Acoust. Soc. Am. 97, 1817–1829.
Lentz, J.J., Leek, M.R., 2001. Psychophysical estimates of cochlear phase response: masking by harmonic complexes. J. Assoc. Res. Otolaryngol. 2, 408–422.
Levitt, H., 1971. Transformed up-down methods in psychophysics. J. Acoust. Soc. Am. 49, 467–477.
Mao, J., Vosoughi, A., Carney, L.H., 2013. Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. J. Acoust. Soc. Am. 134, 396–406.
Mao, J., Carney, L.H., 2014. Binaural detection with narrowband and wideband reproducible noise maskers. IV. Models using interaural time, level, and envelope differences. J. Acoust. Soc. Am. 135, 824–837.
Mao, J., Carney, L.H., 2015. Tone-in-noise detection using envelope cues: comparison of signal-processing-based and physiological models. J. Assoc. Res. Otolaryngol. 16, 121–133.
Maxwell, B.N., Richards, V.M., Carney, L.H., 2020. Neural fluctuation cues for simultaneous notched-noise masking and profile-analysis tasks: insights from model midbrain responses. J. Acoust. Soc. Am. 147, 3523–3537.
Micheyl, C., Maison, S., Carlyon, R.P., Andéol, G., Collet, L., 1999. Contralateral suppression of transiently evoked otoacoustic emissions by harmonic complex tones in humans. J. Acoust. Soc. Am. 105, 293–305.
Mishra, S.K., Biswal, M., 2019. Neural encoding of amplitude modulations in the human efferent system. J. Assoc. Res. Otolaryngol. 20, 383–393.
Mehrgardt, S., and Schroeder, M.R. (1983). Monaural phase effects in masking with multicomponent signals, In: Klinke R., and Hartmann R. (Eds) Hearing Physiological Bases and Psychophysics (Springer, Berlin). https://doi.org/10.1007/978-3-642-69257-4_42
Münkner, S., Kohlrausch, A.G., Püschel, D., 1996. Influence of fine structure and envelope variability on gap-duration discrimination thresholds. J. Acoust. Soc. Am. 99, 3126–3137.
Nelson, P.C., Carney, L.H., 2004. A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J. Acoust. Soc. Am. 116, 2173–2186.
Oxenham, A.J., Moore, B.C., Vickers, D.A., 1997. Short-term temporal integration: evidence for the influence of peripheral compression. J. Acoust. Soc. Am. 101, 3676–3687.
Oxenham, A.J., Dau, T., 2001a. Reconciling frequency selectivity and phase effects in masking. J. Acoust. Soc. Am. 110, 1525–1538.
Oxenham, A.J., Dau, T., 2001b. Towards a measure of auditory-filter phase response. J. Acoust. Soc. Am. 110, 3169–3178.
Oxenham, A.J., Dau, T., 2004. Masker phase effects in normal-hearing and hearing-impaired listeners: evidence for peripheral compression at low signal frequencies. J. Acoust. Soc. Am. 116, 2248–2257.
Plack, C.J., 2018. The Sense of Hearing. Routledge Press.
Plomp, R., Bouman, M.A., 1959. Relation between hearing threshold and duration for tone pulses. J. Acoust. Soc. Am. 31, 749–758.
Pressnitzer, D., Meddis, R., Delahaye, R., Winter, I.M., 2001. Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J. Neurosci. 21, 6377–6386.
Richards, V.M., 1992. The detectability of a tone added to narrow bands of equal-energy noise. J. Acoust. Soc. Am. 91, 3424–3435.
Schooneveldt, G.P., Moore, B.C., 1989. Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration. J. Acoust. Soc. Am. 85, 273–281.
Schroeder, M., 1970. Synthesis of low peak-factor signals and binary sequences of low autocorrelation. IEEE Trans. Inf. Theory 16, 85–89.
Shen, Y., Lentz, J.J., 2009. Level dependence in behavioral measurements of auditory-filter phase characteristics. J. Acoust. Soc. Am. 126, 2501–2510.
Shen, Y., Pearson, D.V., 2017. Recognition of synthesized vowel sequences in steady-state and sinusoidally amplitude-modulated noises. J. Acoust. Soc. Am. 141, 1835–1841.
Shera, C.A., Guinan, J.J., Oxenham, A.J., 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc. Natl. Acad. Sci. U. S. A. 99, 3318–3323.
Smith, B.K., Sieben, U.K., Kohlrausch, A.G., Schroeder, M.R., 1986. Phase effects in masking related to dispersion in the inner ear. J. Acoust. Soc. Am. 80, 1631–1637.
Strickland, E.A., 2001. The relationship between frequency selectivity and overshoot. J. Acoust. Soc. Am. 109, 2062–2073.
Strickland, E.A., Viemeister, N.F., 1996. Cues for discrimination of envelopes. J. Acoust. Soc. Am. 99, 3638–3646.
Summers, V., 2000. Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes. J. Acoust. Soc. Am. 108, 2307–2317.
Summers, V., Leek, M.R., 1998. Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners. Hear. Res. 118, 139–150.
Tabuchi, H., Laback, B., Necciari, T., Majdak, P., 2016. The role of compression in the simultaneous masker phase effect. J. Acoust. Soc. Am. 140, 2680–2694.
Tabuchi, H., Laback, B., 2017. Psychophysical and modeling approaches towards determining the cochlear phase response based on interaural time differences. J. Acoust. Soc. Am. 141, 4314–4331.
Tabuchi, H., Laback, B., 2020. The roles of long-term envelope regularity and efferent activation in the simultaneous masker phase effect. In: The 43rd Association for Research in Otolaryngology (ARO) Meeting. San Jose, California.
Viemeister, N.F., Wakefield, G.H., 1991. Temporal integration and multiple looks. J. Acoust. Soc. Am. 90, 858–865.
Walsh, K.P., Pasanen, E.G., McFadden, D., 2010. Overshoot measured physiologically and psychophysically in the same human ears. Hear. Res. 268, 22–37.
Wojtczak, M., Schroder, A.C., Kong, Y.Y., Nelson, D.A., 2001. The effect of basilar-membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing. J. Acoust. Soc. Am. 109, 1571–1586.
Wojtczak, M., Oxenham, A.J., 2009. On- and off-frequency forward masking by Schroeder-phase complexes. J. Assoc. Res. Otolaryngol. 10, 595–607.
Wojtczak, M., Beim, J.A., Oxenham, A.J., 2015. Exploring the role of feedback-based auditory reflexes in forward masking by Schroeder-phase complexes. J. Assoc. Res. Otolaryngol. 16, 81–99.
Yasin, I., Drga, V., Plack, C.J., 2014. Effect of human auditory efferent feedback on cochlear gain and compression. J. Neurosci. 34, 15319–15326.
Zilany, M.S., Bruce, I.C., Carney, L.H., 2014. Updated parameters and expanded simulation options for a model of the auditory periphery. J. Acoust. Soc. Am. 135, 283–286.
Zilany, M.S., Bruce, I.C., Nelson, P.C., Carney, L.H., 2009. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. J. Acoust. Soc. Am. 126, 2390–2412.