The role of compression in the simultaneous masker phase effect[a)]

Hisaaki Tabuchi, Bernhard Laback,[b)] Thibaud Necciari, and Piotr Majdak
Austrian Academy of Sciences, Acoustics Research Institute, Wohllebengasse 12-14, 1040 Vienna, Austria

(Received 19 November 2015; revised 18 July 2016; accepted 5 September 2016; published online 17 October 2016)

Peripheral compression is believed to play a major role in the masker phase effect (MPE). While compression is almost instantaneous, activation of the efferent system reduces compression in a temporally evolving manner. To study the role of efferent-controlled compression in the MPE, in experiment 1, simultaneous masking of a 30-ms 4-kHz tone by 40-ms Schroeder-phase harmonic complexes was measured with on- and off-frequency precursors as a function of masker phase curvature for two masker levels (60 and 90 dB sound pressure level). The MPE was quantified by the threshold range [min/max difference (MMD)] across the phase curvatures. For the 60-dB condition, the presence of on-frequency precursor decreased the MMD from 10 to 5 dB. Experiment 2 studied the role of the precursor on the auditory filter's bandwidth. The on-frequency precursor was found to increase the bandwidth, an effect incorporated in the subsequent modeling. A model of the auditory periphery including cochlear filtering and basilar membrane compression generally underestimated the MMDs. A model based on two-step compression, including compression of inner hair cells, accounted for the MMDs across precursor and level conditions. Overall, the observed precursor effects and the model predictions suggest an important role of compression in the simultaneous MPE. © *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4964328]

[EAS] Pages: 2680–2694

I. INTRODUCTION

Over the last three decades, many studies assessed the phase response of the human auditory system by means of a masking paradigm involving Schroeder-phase harmonic complexes (Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a,b; Summers and Leek, 1998; Leek *et al.*, 2000; Summers, 2000, 2001; Oxenham and Dau, 2001a,b, 2004; Lentz and Leek, 2001; Oxenham and Ewert, 2005; Rupp *et al.*, 2008; Shen and Lentz, 2009; Wojtczak and Oxenham, 2009; Wojtczak *et al.*, 2015). Nonetheless, the mechanisms underlying measures of the auditory phase response are not fully understood yet. The present work investigates the particular role of peripheral compression in the estimate of the phase response of auditory filters (AFs) based on a simultaneous masking paradigm.

The phase response of AFs is usually estimated by measuring the amount of masking of a target tone evoked by a Schroeder-phase harmonic-complex masker (Schroeder, 1970). The phase curvature of a Schroeder-phase stimulus is defined as the second derivative of the unwrapped[1] phase as a function of frequency (Kohlrausch and Sander, 1995),

$$\frac{d^2\theta}{df^2} = C\frac{2\pi}{(N_2 - N_1 + 1)f_0{}^2}, \tag{1}$$

where C determines the phase curvature, f_0 is the fundamental frequency, and N_1 and N_2 are the lowest and highest harmonics, respectively. Given that N_1, N_2, and f_0 are fixed, the masker phase curvature depends only on C. One important property of Schroeder-phase stimuli is that their crest factor, a measure of the peakiness of the waveform, changes with C, ranging from very peaky envelopes for $C = 0$ to flat envelopes for $C = -1$ or 1. When a Schroeder-phase signal is filtered by the AF, the phase response of the AF is added to the phases of the individual signal components, resulting in a more or less peaky "internal" signal representation depending on C (Kohlrausch and Sander, 1995). The internal signal representation is thus assumed to be maximally peaky when the phases of the individual signal components correspond to the inverse of the AF's phase response (or, in other terms, when signal's C is the inverse of the AF's C).

The amount of masking depends on the masker's phase curvature, an effect referred to as the masker phase effect (MPE). A maximally peaky masker's internal representation produces the smallest masking effect (Kohlrausch and Sander, 1995). There are two main explanations for this effect. First, listeners may detect the target at instances where the internal masker representation has a temporal dip (Kohlrausch and Sander, 1995). The more pronounced the dips the easier to detect the target in the dips. However, the MPE has been found in both forward and simultaneous masking configurations, which is difficult to be explained simply in terms of dip listening (Carlyon and Datta, 1997a). Second, the instantaneous compression of the masker may result in lower excitation levels and in turn yield less masking for peaky than for flat maskers (Carlyon and Datta, 1997a).

The most important physiological entities responsible for peripheral compression are the outer hair cells (OHCs)

[a)]Portions of this work were presented at the 37th ARO meeting in San Diego, 2014.
[b)]Electronic mail: Bernhard.Laback@oeaw.ac.at

on the basilar membrane (BM) in the cochlea. OHCs amplify incoming sounds within the order of less than a millisecond (e.g., Recio *et al.*, 1998). The quasi-instantaneous amplification by the OHCs particularly operates at low to moderate signal amplitudes (e.g., Ruggero and Rich, 1991). The input-output (I/O) characteristic of the inner hair cells (IHCs) is also somewhat compressive, particularly at high input levels (e.g., Russell and Sellick, 1978; Dallos, 1985; Zhang *et al.*, 2001; Lopez-Poveda and Eustaquio-Martín, 2006). One important property of the gain provided by the OHCs is that it is modulated by means of efferent feedback via the medial olivocochlear (MOC) system (e.g., Guinan, 2006). In particular, activation of the MOC system is known to reduce that gain in a frequency-specific, temporally evolving, and level-dependent manner (Lilaonitkul and Guinan, 2009). In humans, the latency between the onset of MOC system activation and the onset of gain reduction is in the order of 30 ms (Backus and Guinan, 2006; Roverud and Strickland, 2010). The MOC-induced gain reduction appears to be strong enough to linearize the I/O function of the BM, i.e., to reduce compression (e.g., von Klitzing and Kohlrausch, 1994; Krull and Strickland, 2008; Laback *et al.*, 2011; Yasin *et al.*, 2014).

These properties of the MOC system provide an opportunity to study the impact of cochlear compression on the MPE in normal-hearing listeners. Our approach was to measure the MPE with and without the MOC system activated by a precursor. The masker duration was kept short (40 ms) aiming at minimizing the MOC system activation. Thus, conditions without the precursor were presumed to reflect the effect of full gain and compression, and thus result in the greatest MPE. In contrast, conditions with the precursor were presumed to reflect reduction of that gain and compression, and thus result in a smaller MPE.

Listeners with cochlear hearing impairment, who generally exhibit reduced compression and poor frequency selectivity, show a much smaller MPE than normal-hearing listeners (Summers and Leek, 1998; Summers, 2000, 2001; Oxenham and Dau, 2004). This is consistent with the idea that compression is important for the MPE. On the other hand, widening of AFs as a consequence of reduced OHC gain may actually *increase* the MPE because of the increased number of components interacting within a given AF (Oxenham and Dau, 2004). A similar effect of AF widening occurs in normal-hearing listeners at high masker levels (e.g., Glasberg and Moore, 1990). The observation of low MPE in hearing-impaired listeners seems to suggest that the amount of compression is more important than the AF bandwidth for the MPE. In order to better control for and quantify the influence of spectral filtering on the MPE, in our study, we intentionally limited the masker bandwidth and measured the actual bandwidth of the AFs with and without a precursor.

The organization of this paper is as follows. Experiment 1 investigated the influence of a precursor on the MPE for two masker levels. An on-frequency precursor was expected to reduce compression, resulting in reduced MPE. In experiment 2, the bandwidth of AFs for signal conditions similar to experiment 1 (including precursor) was measured using a notched-noise paradigm, in order to estimate the effect of AF bandwidth on the MPE. Last, a model of peripheral auditory processing is proposed to predict the MPE from experiment 1, based on the AF shapes obtained in experiment 2.

II. EXPERIMENT 1: MASKER PHASE EFFECT WITH AND WITHOUT PRECURSOR

A. Listeners and equipment

Eight subjects aged between 19 and 33 years participated in the experiment. All had absolute hearing thresholds of 20 dB (*re* 20 μPa) or lower at octave frequencies between 0.25 and 8 kHz. Six of the listeners had experience from previous psychophysical experiments. Listeners received monetary compensation for their participation. None of the authors participated in the experiment.

The experiments were conducted in a double-walled, sound attenuating booth. The stimuli were output via a 24-bit A/D-D/A converter (AD/DA 2402, Digital Audio Denmark) using a sampling rate of 48 kHz. The analog signals were sent through a headphone amplifier (HB6, Tucker-Davis Technologies, TDT, Alachua, FL), and an attenuator (PA4, TDT) and finally fed to circumaural headphones (HDA200, Sennheiser). Calibration of the stimuli (separately done for precursor, masker, and target) was performed using a sound level meter (2260, Brüel & Kjær) connected to an artificial ear (4153, Brüel & Kjær).

B. Stimuli

We chose the simultaneous masking paradigm for consistency and better comparability with a number of relevant studies from the literature (e.g., Oxenham and Dau, 2004). The masker was a Schroeder-phase harmonic complex (Schroeder, 1970; Lentz and Leek, 2001) defined as

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

with $f_0 = 100$ Hz because of strong monaural MPE observed at this frequency (Smith *et al.*, 1986). We used $N_1 = 34$ and $N_2 = 46$ yielding a narrow masker bandwidth (3400 to 4600 Hz) in order to (1) reduce the potentially confounding effect of spectral filtering and (2) reduce the potential effect of masker components outside of the AF under test (i.e., off-frequency) whose phase curvature is different from that of the on-frequency components (Oxenham and Ewert, 2005) and thus likely violates the assumption of a constant phase curvature underlying the masking paradigm. The masker had a duration of 40 ms.

The target was a 30-ms, 4000-Hz pure tone temporally centered at the masker. Consequently, the masker-onset to target-onset interval was 5 ms, which is shorter than the MOC latency. Therefore, no masker-related MOC-induced gain reduction was expected during the target presentation. C's ranged from -1 to 1 in intervals of 0.25. The target was added in cosine phase to the 4000-Hz masker component. The nominal target amplitude was then "corrected" for all nonzero C's to account for phase interactions between the target and the 4000-Hz masker component.

The precursor was a 400-ms sinusoid preceding the masker without any gap between precursor offset and masker onset. This duration was intended to activate the MOC system (Backus and Guinan, 2006). Two precursor frequencies were tested: 4000 Hz (on-frequency) and 800 Hz (off-frequency). In addition, there was a control condition without precursor (no-precursor). The off-frequency precursor condition was intended to disentangle the effects of the precursor *per se* and other effects related to the presence of a preceding sound (e.g., cueing or grouping). Being more than two octaves below the target frequency, the off-frequency precursor was expected to yield almost the same thresholds as without any precursor (Jennings *et al.*, 2009).

Two overall sound pressure levels (SPLs) were tested: 60 and 90 dB. Precursor and masker were presented at the same SPL because the precursor effect tends to be largest when the precursor SPL equals the masker SPL (Bacon and Healy, 2000). The masker, target, and precursor were gated on and off with 5-ms cosine-squared ramps.

Continuous background noise was added to mask low-frequency 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). The overall SPL of the continuous noise was 55 and 70 dB for the 60- and 90-dB level conditions, respectively. Stimuli were presented to the right ear.

C. Procedure

An adaptive three-interval forced choice procedure with a three-down one-up staircase rule was used to measure thresholds at 79% correct (Levitt, 1971). The intervals between the three stimuli of a trial were 500 ms, which allowed enough time for the efferent system to recover (Walsh *et al.*, 2010). 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 eight reversals. The step size was 4 dB for the first four reversals, and then reduced to 2 dB for the last four reversals. The target threshold was estimated from the average of the last four reversals. One threshold estimate took about five minutes.

Listeners were tested in blocks. Each block contained one of the two masker levels and all combinations of the nine C's and the three precursor conditions, resulting in 27 thresholds. Each block lasted for approximately 2.5 h. Listeners were allowed to take a break every three runs in a block.

In the experiment, each block was tested three times for each of the masker levels, resulting in six blocks tested. The level at the first block was randomly chosen for each listener and alternated across the listener's block repetitions. The total testing time of the experiment amounted to about 15 to 18 h per listener. We visually checked the masked thresholds as a function of the number of blocks, and did not observe any learning or training effects.

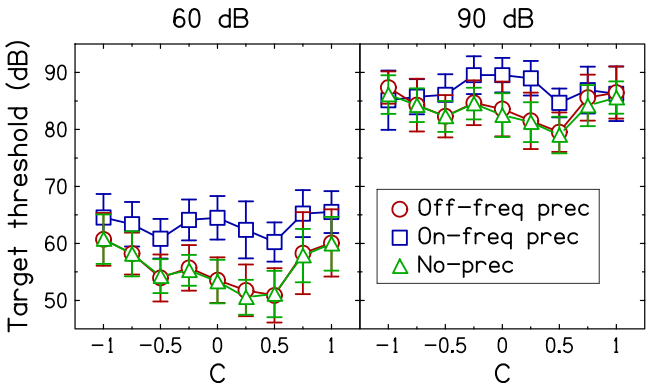

FIG. 1. (Color online) Mean target threshold as a function of C. The left and right panels show the results for various precursor conditions when the presentation level of masker and precursor was 60 and 90 dB SPL, respectively. Error bars indicate ±1 standard deviation across listeners.

D. Results

Figure 1 shows mean target thresholds across listeners as a function of C with the precursor condition as the parameter. The results for the 60- and 90-dB conditions are shown in the left and right panels, respectively. Thresholds in the off-frequency (circles) and no-precursor (triangles) conditions were very similar, both showing a minimum in the vicinity of $C = 0.5$. Thresholds in the on-frequency precursor condition (squares) were significantly higher and more similar across C's than those in the off-frequency and no-precursor conditions.

The threshold range across the tested C's quantifies the MPE and is referred to as min/max difference (MMD). The MMDs for individual listeners are shown in Table I. The MMDs, averaged across the listeners, decreased from about 10 dB (in the no- or off-frequency precursor conditions) to 5.3 dB (on-frequency conditions) in the 60-dB condition, and decreased from 7.5 to 4.9 dB in the 90-dB condition. A three-way repeated-measures analysis of variance (ANOVA) showed main effects of each factor; phase curvature C [$F(8,56) = 49.23$, $p < 0.001$], precursor [$F(2,14) = 58.48$, $p < 0.001$], and presentation level [$F(1,7) = 2351$, $p < 0.001$]. The different shapes of the threshold functions are reflected by the significant interaction between

TABLE I. Differences between the maximum and minimum target thresholds in dB (min/max difference; MMD) for different precursor and level conditions obtained in experiment 1. The bottom row shows the MMD obtained from the averaged thresholds across all the listeners.

	60 dB			90 dB		
Listener	No-prec.	Off-freq.	On-freq.	No-prec.	Off-freq.	On-freq.
NH39	14.8	12.8	5.4	7.2	7.3	7.4
NH43	11.3	11.6	11.7	7.8	11.6	4.2
NH47	17.7	16.6	5.3	7.4	7.8	6.2
NH84	9.8	9.6	7.1	6.7	9.9	7.4
NH136	10.2	10.2	5.6	5.6	6.1	8.3
NH143	14.2	15.1	7.1	9.8	9.4	9.0
NH144	8.5	6.3	5.4	11.9	12.3	4.9
NH145	7.4	5.8	5.7	10.6	8.8	6.2
Mean	10.2	9.8	5.3	7.1	7.8	4.9

the factors C and precursor [$F(16,112) = 20.03$, $p < 0.001$]. The three-way interaction between C, precursor (the data for off-frequency and no-precursor), and presentation level was not significant [$F(8,56) = 0.75$, $p = 0.64$], suggesting that MMDs did not substantially change with the presentation levels. A separate two-way repeated-measures ANOVA showed no significant differences between the off-frequency and no-precursor conditions both at 60 dB [$F(1,7) = 0.14$, $p = 0.71$] and 90 dB [$F(1,7) = 2.61$, $p = 0.15$]. *Post hoc* pairwise comparisons with the Tukey LSD test indicated that for both levels the thresholds of off-frequency precursor and no precursor conditions differed significantly between $C = 0.5$ and all other C's [$p \leq 0.02$] with the only exception of the difference between $C = 0.5$ and 0.25 at 60 dB [$p \geq 0.21$].

E. Discussion

Overall, the on-frequency precursor raised thresholds for all C's and significantly decreased the MMDs. These results are consistent with a reduction in gain and thus reduction of compression at the target frequency. This gain reduction may be attributed to the activation of the efferent system by the on-frequency precursor. Note that the precursor might also have contributed to the smaller effect of C by its role as a forward masker, affecting the role of the simultaneous masker in the simultaneous masking paradigm. Although we assumed this effect to be small, we have checked its contribution by means of modeling the nonlinear additivity of masking, based on psychophysically estimated compression in humans (e.g., Plack *et al.*, 2008; Laback *et al.*, 2011), resulting from the two maskers (simultaneous masker, and precursor). The results (for more details see the Appendix) suggest that while forward masking by the precursor may have contributed somewhat to the reduced effect of C, particularly for the 60-dB masker conditions, it does not explain the experimentally observed effects, favoring the idea of the precursor reducing the cochlear compression by means of efferent control (see also Roverud and Strickland, 2014).

In the no-precursor condition, our MMDs for the 90-dB masker were clearly smaller than those in the order of 20 dB reported in studies using comparable masker levels (e.g., Lentz and Leek, 2001; Oxenham and Dau, 2001b; Shen and Lentz, 2009). Generally, MMDs seem to decrease with decreasing masker frequency range (Oxenham and Dau, 2001a). Thus, our smaller MMDs may be due to the much smaller frequency range of 3400 to 4600 Hz, as compared to others' ranges of 200 to 5000 Hz (Lentz and Leek, 2001) and 1600 to 6400 Hz (Oxenham and Dau, 2001b; Shen and Lentz, 2009). We addressed that issue by measuring masked thresholds for the no-precursor condition in a control experiment with four of the listeners (NH39, NH43, NH143, NH144) using a masker SPL of 90 dB (as in Oxenham and Dau, 2001b; Shen and Lentz, 2009). For maskers with a bandwidth in the range 1600 to 6400 Hz (as in Oxenham and Dau, 2001b) and a duration of 40 ms (as in experiment 1), the mean MMD was 8.3 dB. This is similar to the mean MMD of 7.7 dB from our experiment 1 for those listeners. For the wider masker bandwidth (as in Oxenham and Dau, 2001b) in combination with longer durations of *both* masker and target (320 and 310 ms, respectively, as in Oxenham and Dau, 2001b) the mean MMD was 16.9 dB, much closer to the MMD reported in Oxenham and Dau (2001b). One potential explanation for the larger MPE observed for the long masker *and* target could be that a long-duration target facilitates target detection in the multiple dips of a peaky masker compared to a short-duration target, thus reducing thresholds. For example, when considering a "multiple looks" type of temporal integration (Viemeister and Wakefield, 1991), the target information within individual envelope dips may be integrated across multiple dips, ignoring the envelope peaks in between. This would result in low thresholds particularly for peaky maskers in combination with a long masker and target duration. Further examination of the mechanisms underlying these effects is, however, beyond the scope of this study.

In two previous studies, the MMDs increased with the masker level (Oxenham and Dau, 2001b; Shen and Lentz, 2009), whereas the present study did not show such a level dependency (if at all, we observed smaller MMDs at the higher level). The greater MMDs at high levels (around 90 dB) compared to lower levels reported in those studies appears to be counter-intuitive because the BM I/O function at a characteristic frequency (CF) has been found to be either constantly compressive up to about 90 dB (e.g., Ruggero *et al.*, 1997) or approach linearity around 90 dB (e.g., Yasin *et al.*, 2014). Such a reduced compression (as compared to that at lower levels) is supposed to lead to a decreasing (rather than increasing) MPE. Other studies actually showed a reduction of the MMDs as the masker level increased from mid to high levels (Summers and Leek, 1998; Summers, 2000; Wojtczak and Oxenham, 2009). As noticed by Summers *et al.* (2003) and Shen and Lentz (2009), the expanding AF bandwidth with increasing level causes the AF phase curvature to approach zero for low-(off)-frequency components (see Oxenham and Ewert, 2005). Thus, the change in MMDs as an effect of changing masker level might reflect the effect of altering the number of components interacting within the effective AF rather than the change of AF phase curvature *per se*. We attempted to avoid such confounding effects by using narrowband maskers for which even the lowest components did not fall within the off-frequency region half an octave or more below CF where the phase-curvature approaches zero (Shera, 2001).

III. EXPERIMENT 2: THE AF BANDWIDTH

In experiment 1, the MPE was found to be reduced by presenting an on-frequency precursor. We suggested that this could be attributable to a reduction of cochlear gain as a consequence of efferents' activation. Another consequence of reduced cochlear gain may be the increase in AF bandwidth, as shown by comparing (1) listeners with normal hearing and cochlear hearing loss (e.g., Glasberg and Moore, 1986), (2) AF bandwidth differences in normal-hearing listeners between conditions with and without precursor (e.g., Strickland, 2001; Jennings and Strickland, 2012), and (3) physiological measurements of AF filter tuning with and without deactivation of OHC (Ruggero and Rich, 1991) and MOC activity (Walsh *et al.*, 1998). In experiment 2, we thus

aimed to quantify the amount of AF bandwidth change caused by the presence of the precursor in experiment 1. To that end, we attempted to estimate the bandwidth of the AF centered on the target frequency in conditions with on- and off-frequency precursors. The precursor and target stimuli and the temporal stimulus configurations were the same as in experiment 1. The filter shapes were estimated using the simultaneous masking paradigm.

A. Stimuli and procedure

We used the notched-noise method to estimate the filter's shape (Patterson, 1976). This method assumes that the auditory system integrates the power spectrum over frequency and disregards the signal's temporal properties. For the masker, we used noises having the same duration and bandwidth as the maskers from experiment 1. The notched-noise was produced by combining two narrow-band noises with a bandwidth of 600 Hz and slopes of about 60 dB/oct each. Thus, the masker bandwidth amounted to only 15% of the target frequency. Each noise band was generated by filtering Gaussian white noise with 1024th-order finite impulse response filters (i.e., the length of the impulse response was 1025 samples). Normalized notch width was defined as the frequency distance between the target frequency and the upper edge of the lower-frequency band, and between the target frequency and the lower edge of high-frequency band, divided by the target frequency. We tested symmetrical normalized notch widths of 0.0 (a single band-pass with 1200-Hz bandwidth), 0.1, 0.2, 0.4, 0.6, and 0.8, as well as asymmetrical notch widths of (0.2, 0.4 for the lower and upper band, respectively), (0.4, 0.2), (0.4, 0.6), and (0.6, 0.4). The same noise token was used for the three intervals of a trial ("frozen" noise), but new noise tokens were used across trials. The intent was to reduce the variability in listeners' performance which may be caused by the random spectral variation across trials (Rosen and Baker, 1994).

Besides the use of notched noise stimuli, the following aspects of the methodology changed compared to experiment 1: (1) the no-precursor condition was tested only for the notch width of zero because of similar results in the no-precursor and off-frequency precursor conditions obtained in experiment 1, (2) conditions without notched noise (no-noise) were included corresponding to the measurement of tone threshold in quiet in the no-precursor condition and to the measurement of forward masking in precursor conditions, (3) the low-frequency, continuous noise was not used because the effect of distortion products was not considered as being critical in this paradigm (e.g., Patterson *et al.*, 1982), (4) listener NH136 was no longer available.

B. Results and discussion

Figure 2 shows target thresholds averaged across listeners as a function of the normalized notch width. Left and right panels show results for the 60- and 90-dB levels, respectively. The thresholds for the asymmetrical notches are shown at the notch width corresponding to the smaller of the two notch widths.

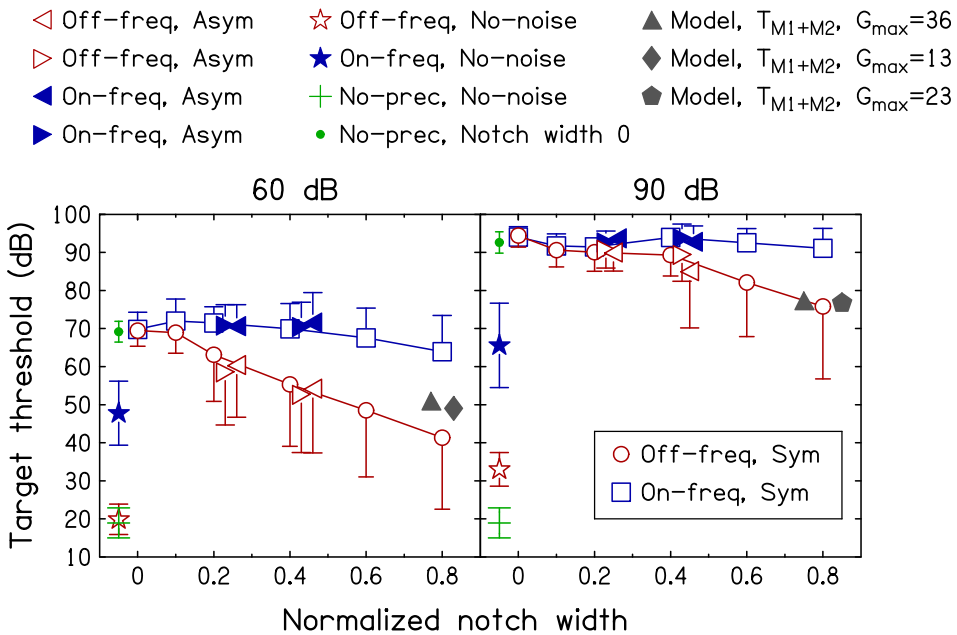

FIG. 2. (Color online) Mean target threshold as a function of normalized notch width. The left and right panels show the results for precursor conditions when the SPL of notched-noise maskers and precursor was 60 and 90 dB SPL, respectively. For clarity, some error bars denote either plus or minus one standard deviation across listeners and some data points are slightly shifted along the horizontal axis. The symmetrical off- and on-frequency conditions are denoted by the symbol in the inset and connected by a line. The symbols for the other conditions are shown above the figure panels. As for the asymmetric condition, the normalized notch width corresponds to the smaller notch width. The larger notch width toward *lower* frequencies is denoted by the *left-pointing* triangles, whereas the larger notch width toward *higher* frequencies is denoted by the *right-pointing* triangles. On the left side of each panel, the target thresholds without notched-noise (no-noise) and for a notch width of zero (band-pass noise) without precursor are shown. Target thresholds (in dB) predicted by the masking-additivity model are shown at the normalized notch width of 0.8 (slightly shifted along the horizontal axis for clarity). T_{M1+M2} denotes the predicted target threshold in presence of both the precursor (M1) and the notched-noise masker (M2). The triangle symbols show predictions assuming full target compression; the diamonds and pentagons show predictions assuming reduced compression (for details, see the Appendix).

The thresholds were largest for the normalized notch width of zero and decreased, more or less depending on the condition, as the notch width increased. Compared to the off-frequency precursor condition, the thresholds in the on-frequency precursor condition were increasingly elevated with increasing notch width, resulting in largely flattened threshold patterns. Also, increasing the level from 60 to 90 dB SPL flattened the threshold patterns. A three-way repeated-measures ANOVA showed significant main effects of each factor; precursor [$F(1,6) = 15.71$, $p < 0.01$], presentation level [$F(1,6) = 228.6$, $p < 0.001$], and notch width [$F(5,30) = 9.83$, $p < 0.001$]. Differences in the slopes of the threshold functions are reflected by the significant interactions between the factors notch width and precursor [$F(5,30) = 15.28$, $p < 0.001$] and between the factors notch width and presentation level [$F(5,30) = 6.34$, $p < 0.001$], suggesting that the AF bandwidth increased significantly both with increasing level and adding an on-frequency precursor.

In Fig. 2, the thresholds obtained for the no-noise conditions are shown at the far left of each panel. The threshold difference between *on-frequency precursor* and *no-precursor* conditions represents the effect of the precursor. This difference was considerable, with average values of approximately 29 and 48 dB in the 60- and 90-dB level conditions, respectively. The threshold difference between *off-frequency precursor* and *no-precursor* conditions was much smaller, with averages of approximately 1 and 14 dB in the two level conditions, respectively, and most probably negligible compared to the thresholds obtained in the presence of the masker.

As in experiment 1, the precursor might have acted as forward masker interacting with the masking of the simultaneous maskers. We thus checked the additivity of masking resulting from the two maskers (notched noise and precursor), using a condition for which masking additivity could be simulated in a straight-forward manner. The results (for more details see the Appendix) indicate that, despite a small contribution of the additivity, the elevation of notched-noise thresholds by the on-frequency precursor does not appear to be explainable by the additivity of masking based on psychophysically estimated compression in humans. The results appear to be more consistent with the idea of precursor-induced gain reduction (e.g., Strickland, 2001; Jennings and Strickland, 2012).

Similar thresholds were obtained for the symmetrical and asymmetrical notch widths. This is in agreement with findings showing symmetrical AF shapes when tested with narrow noise bands as in the present study (Lutfi and Patterson, 1984) and asymmetrical AF shapes when tested with wider noise bands (at least 40% of the target frequency, e.g., Patterson *et al.*, 1982; Glasberg and Moore, 1990).

C. Modeling of AF shapes

The rounded exponential (roex) model (Patterson *et al.*, 1982) is assumed to represent the shape of the AFs. In this model, the AF slopes for the lower and upper skirts are represented by the parameters p_l and p_u, respectively. For the normalized frequency $g = |f - f_c|/f_c$, where f_c is the target frequency, the AF shape $W(g)$ is

$$W(g) = (1 + pg)e^{-pg}, \tag{3}$$

with p being p_l and p_u, for the lower and upper slope, respectively. Then, according to the power spectrum model of masking, P, i.e., the power of the target at a threshold obtained from experiment 2, is given by

$$P = K_{AF} \int N_0 W(g) dg, \tag{4}$$

with K_{AF} representing the listener's sensitivity in experiment 2 and N_0 representing the spectrum level of the presented noise (for further details, see Patterson, 1976; Patterson *et al.*, 1982).

In the fitting process, p_l and p_u were freely varied (assuming at first $K_{AF} = 1$) and K_{AF} was taken as the mean deviation between data and predictions across all the normalized notch widths (Patterson *et al.*, 1982). We estimated only these two free parameters to minimize the RMS errors between data and prediction because the fit seemed unstable when K_{AF} was also allowed to vary (for further details on the fitting, see Patterson *et al.*, 1982). Then, the equivalent rectangular bandwidths (ERBs) were derived from the AF shapes represented by the parameters p_l *and* p_u (Glasberg and Moore, 1990). The estimated parameters and their corresponding ERBs for individual listeners are listed in Table II. In some conditions and for some listeners, masked thresholds remained almost constant across the tested notch widths. As a result, the AF slopes were nearly zero and the ERB estimates were "infinite."

Figure 3 shows the AF shapes $W(g)$ estimated from the mean thresholds across listeners ("mean" in Table II). As expected, the AFs became wider both with increasing level and the presence of the on-frequency precursor. In the off-frequency precursor condition, the level increase from 60 to 90 dB raised the ERB from 1466 to 2283 Hz, which corresponds to a factor of 1.6. For the 60-dB level condition, the presence of the on-frequency precursor increased the ERB from 1466 to 5032 Hz, which corresponds to a considerably larger factor, namely, 3.4. These results qualitatively support our conclusions from the masked thresholds that the presence of the on-frequency precursor expands the bandwidth, probably even more severely than the level.

For the no-precursor 60-dB level conditions, the estimated ERB of 1466 Hz was much larger than the ERB of 456 Hz given by the formula ERB = $24.7(4.37F + 1)$, which describes the AF bandwidth as a function of center frequency, F (in kHz) (Glasberg and Moore, 1990). However, the ERB has been later found to be actually wider (Unoki *et al.*, 2006), being between 500 and 900 Hz and depending on the choice of the model architecture (see Unoki *et al.*, 2006). We thus calculated the ERB for the data points for the 4000-Hz target from Glasberg and Moore (2000) using our procedure (see the thick solid line in Fig. 3), obtaining

TABLE II. Parameters of the roex model estimated from the notched-noise masking thresholds of experiment 2 for individual listeners. Mean indicates parameters estimated from the mean thresholds across listeners. Entries labeled as "—" indicate conditions in which the slope steepness approached zero, resulting in infinitely large ERBs.

	60 dB, Off-freq.					90 dB, Off-freq.				
Listener	p_l	p_u	K_{AF} (dB)	rms (dB)	ERB (Hz)	p_l	p_u	K_{AF} (dB)	rms (dB)	ERB (Hz)
NH39	17.5	15.1	10.1	4.0	987	10.5	11.2	9.1	4.7	1472
NH43	24.3	13.5	0.7	6.1	920	17.7	26.6	10.8	6.1	752
NH47	3.6	3.7	19.3	2.0	4397	4.4	1.5	9.6	1.7	7273
NH84	11.4	7.8	11.5	2.4	1733	5.0	4.3	2.2	1.3	3468
NH143	8.4	7.7	20.1	5.3	1980	3.8	5.1	7.7	2.5	3689
NH144	13.7	7.1	3.8	6.1	1705	—	—	—	—	—
NH145	13.8	16.3	8.3	4.2	1069	4.7	5.4	−0.5	2.6	3177
Mean	12.4	9.7	10.2	1.2	1466	6.1	8.3	5.3	1.9	2283
	60 dB, On-freq.					90 dB, On-freq.				
Listener	p_l	p_u	K_{AF} (dB)	rms (dB)	ERB (Hz)	p_l	p_u	K_{AF} (dB)	rms (dB)	ERB (Hz)
NH39	5.9	6.4	11.5	3.5	2591	2.8	3.6	4.4	2.7	5074
NH43	6.4	11.3	13.8	2.1	1951	3.4	4.6	2.3	2.4	4118
NH47	—	—	—	—	—	—	—	—	—	—
NH84	1.9	2.1	11.3	1.5	8019	—	—	—	—	—
NH143	3.7	3.8	20.2	2.6	4302	—	—	—	—	—
NH144	—	—	—	—	—	—	—	—	—	—
NH145	—	—	—	—	—	—	—	—	—	—
Mean	3.2	3.2	12.5	1.5	5032	—	—	—	—	—

an ERB of 738 Hz. While this ERB is still smaller than the ERB estimated based on results from experiment 2, this may be at least partly explained by our shorter stimulus duration (see also Hant *et al.*, 1998; Bacon *et al.*, 2002). A potential explanation for the wider ERB for short maskers may be the lack of MOC involvement which in turn may result in stronger suppression effects as compared to longer maskers (Hegland and Strickland, 2016).

IV. MODEL

Previous models attempting to predict the MPE considered both the effects of auditory filtering and instantaneous compression in the cochlea (Carlyon and Datta, 1997a; Wojtczak and Oxenham, 2009). Carlyon and Datta (1997a) used power-law compression with various exponents in order to predict the MPE in forward masking and discussed the necessity of additional sources of compression besides OHC compression. Wojtczak and Oxenham (2009) also used power-law compression in order to predict their forward masking MPE data for short maskers. Here, we propose a model which follows the basic ideas of previous models, but incorporates novelties in the compression and decision stages in order to predict our results from experiment 1.

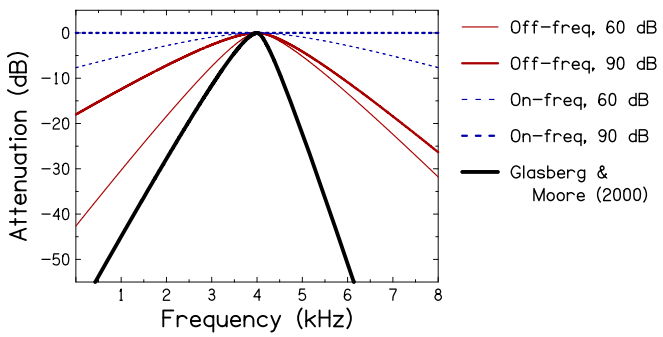

FIG. 3. (Color online) AF shapes [according to Eq. (3)] for the various precursor and level conditions in experiment 2: 60-dB off-frequency (thin), 90-dB off-frequency (thick), 60-dB on-frequency (dotted thin), and 90-dB on-frequency (dotted thick). The parameters estimated from the mean thresholds across listeners are shown in Table II. The thickest solid line shows the AF shape calculated from data by Glasberg and Moore (2000) using our fitting procedure.

A. Model structure

Figure 4 shows the general model structure, including stages for auditory filtering, envelope extraction, compression, and decision. The following notations are used in Fig. 4 and in the following: lowercase letters s denote signal amplitudes in Pa, uppercase letters S denote logarithmic amplitudes in dB, integer subscripts indicate the stage in the signal processing chain, and letter subscripts indicate the signal waveform being processed (M: masker alone; M + T: masker plus target). Letters S without the time variable (t) indicate temporal RMS of logarithmic amplitudes (in dB) computed over the signal length.

1. Auditory filtering

In the first stage, the amplitudes of the stimulus spectral components are weighted by the roex filter shapes described by the parameters p_l and p_u estimated in experiment 2. The filter parameters estimated from the averaged data in each condition (see the bottom row in Table II) are used for the frequency weighting in the corresponding conditions. For the 90-dB on-frequency precursor, no frequency weighting is applied. The AF phase response is considered by

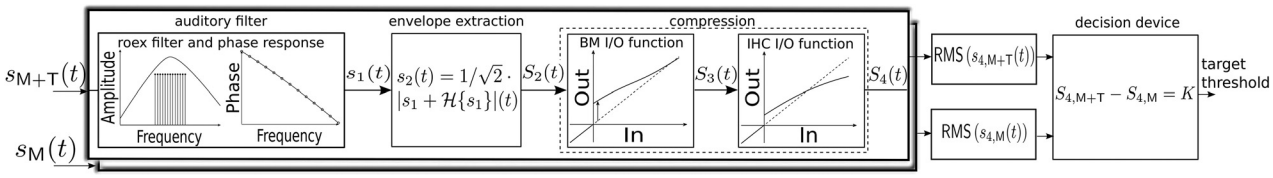

FIG. 4. Schematic structure of the model comprised of four stages: an auditory filter, an envelope extractor, a compressor, and a decision device. Lowercase letters s denote signal amplitudes (in Pa), uppercase S denote logarithmic amplitudes (in dB), integer subscripts indicate the stage in the signal processing chain, and letter subscripts indicate the signal waveform being processed (M: masker alone; M + T: masker plus target). Note that S_2 is obtained from s_2 by taking the logarithm. Letters S without the time variable (t) indicate the logarithmic RMS (in dB) computed over the signal length. For the description of model variables, see text. The dashed lines in the compression stages indicate a linear reference.

subtracting 0.5 from C. The value -0.5 corresponds to the presumed negative AF phase response, given the minimum threshold observed in the results from experiment 1 for $C = 0.5$. This model, thus, implements the spectral weighting (amplitude and phase) already in the stimulus synthesis. For a general model implementation intended for arbitrary input stimuli, the spectral weighting needs to be realized by an appropriate filter.

2. Envelope extraction

In the second stage, the stimulus temporal envelope is calculated by taking the absolute value of the analytic signal (see Fig. 4). A reduction by $\sqrt{2}$ is required because the envelope power is twice the power of the entire waveform including the fine structure (Hartmann, 1998) and because, in the next model stage, the envelope is processed by a BM I/O function based on the entire waveform.

3. Compression

In the third stage, the temporal envelope level $S_2(t)$ is processed in two steps: the first step is intended to mimic instantaneous BM compression, the second step is intended to represent compression by the IHCs. Previous models mimicked BM compression using a power-law function applied to linear signal amplitudes, yielding a constant compression of the signal at all levels (Carlyon and Datta, 1997a; Wojtczak and Oxenham, 2009). We used an I/O function in which the amount of compression depends on signal level (Glasberg and Moore, 2000). This function is thought to be more consistent with both physiological and psychophysical studies (see, e.g., Plack and Arifianto, 2010). Specifically, the BM I/O function is given by Glasberg and Moore (2000),

$$S_3(t) = 0.9S_2(t) + A + B\left[1 - \frac{1}{1 + e^{-0.05(S_2(t)-50)}}\right] \quad \text{for } S_2(t) \geqslant 0,$$
$$S_3(t) = S_2(t) \quad \text{for } S_2(t) < 0, \tag{5}$$

where $A = -0.0894G_{max} + 10.89$, and $B = 1.1789G_{max} - 11.789$. These A and B fix the output $S_3(t)$ at 100 dB for $S_2(t) = 100$ dB (in other terms, the tip of the function is fixed at 100 dB).

G_{max} is the maximum gain applied to $S_2(t)$ of 0 (see also Fig. 6 from Glasberg and Moore, 2000) and is a free parameter to be fit to the actual data. Since the tip of the I/O function is fixed, G_{max} controls the compression of the I/O function. In order to demonstrate this, $S_{3,M}$ was calculated for various G_{max}'s and two C's (-1 and 0.5, corresponding to the maximum and minimum masking conditions, respectively) at the levels of 60 and 90 dB.[2] The result, $S_{3,M}$ versus

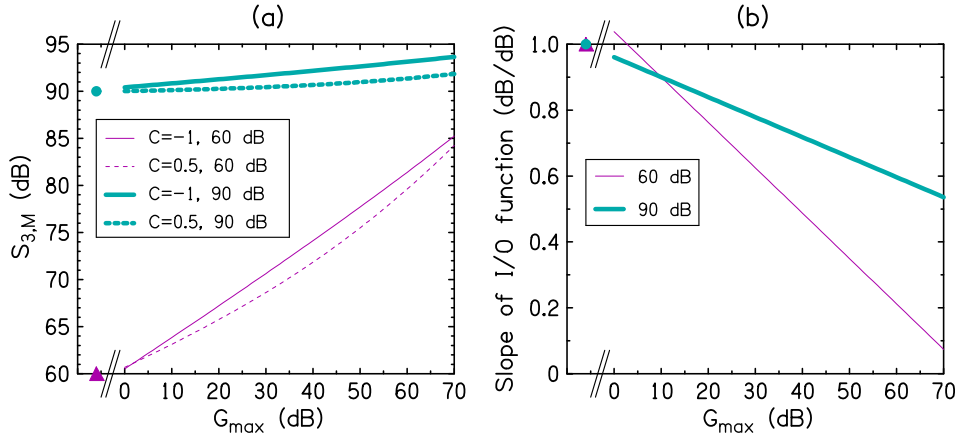

FIG. 5. (Color online) (a): Property of the BM I/O function for peaky ($C = 0.5$) and flat ($C = -1$) maskers, shown as $S_{3,M}$ [in dB, calculated from the RMS value of $s_3(t)$] as a function of G_{max}. The masker waveform $S_M(t)$ was directly used as input $S_2(t)$ (see Fig. 4). Peaky and flat maskers at SPLs of 60 dB (thin) and 90 dB (thick) are shown by dotted and solid lines, respectively. Symbols indicate $S_{3,M}$ when the I/O function has a slope of one. Note that there is no direct link between $S_{3,M}$ shown in (a) and the predicted thresholds shown in Figs. 8 and 9. Here, $S_{3,M}$ was simulated to examine the particular characteristics of BM I/O function in the compression stage. (b): Slopes of the I/O function at 60 (thin) and 90 dB SPL (thick) as a function of G_{max}. The slopes at the 60- and 90-dB input were obtained from the I/O function as given by G_{max}.

G_{max}, is displayed in Fig. 5(a). As expected, the output levels of the peaky waveforms ($C = 0.5$) are smaller than those of the flat waveforms ($C = -1$) for both level conditions. In Fig. 5(a), the symbols indicate the output levels which would be obtained when using a linear I/O function and levels of 60 and 90 dB, respectively. Note that with the linear I/O functions, the output levels remain the same for both C's tested. For the 60-dB level condition, the largest output difference caused by the two curvatures is 2 dB at G_{max} of 50 dB. For the 90-dB level condition, the output differences increase monotonically with G_{max} (within the tested range).

Figure 5(b) shows the corresponding slopes of the I/O functions at the levels of 60 and 90 dB as a function of G_{max}. For 60 dB, the I/O function slope for $G_{max} = 50$ dB, producing the largest output difference in Fig. 5(a), is 0.35 dB/dB.

Within the course of the development of the model, it became clear that the BM I/O function generally underestimates the MPE observed experimentally. Therefore, we conjectured that further compression is required. We extended the compression stage by an additional step mimicking the compression caused by the IHCs (e.g., Russell and Sellick, 1978; Dallos, 1985; Zhang *et al.*, 2001; Lopez-Poveda and Eustaquio-Martín, 2006). We refer to this model as the two-step model. Accordingly, the output of the BM I/O function $S_3(t)$ is passed into an IHC I/O function based on a first-order Boltzmann function,[3]

$$S_4(t) = \frac{1+\beta}{\beta}\left[\frac{\alpha}{1+\beta e^{(Xshift - S_3(t))/\gamma}} - \frac{\alpha}{1+\beta}\right] \quad \text{for } S_3(t) \geq 0,$$
$$S_4(t) = S_3(t) \quad \text{for } S_3(t) < 0. \tag{6}$$

The parameter α limits the maximum output, β and γ determine the slope of the IHC I/O function, and *Xshift* shifts the input to the appropriate level range.

4. Decision

This stage computes a threshold based on the decision variable K representing the listener's sensitivity. The previous model stages were processed twice, once for the masker stimulus alone [$s_M(t)$] and once for the masker-plus-target stimulus [$s_{M+T}(t)$] at the experimentally obtained threshold. K corresponds to the difference between $S_{4,M+T}$ and $S_{4,M}$ (or $S_{3,M+T}$ and $S_{3,M}$ in case of the one-step model). We assume that K is generally invariant within a listener tested across various conditions. Therefore, K was averaged across experimental conditions.

As a quick overview of our general model description, we first determined the decision variable K which best predicted masked thresholds of an "average" listener, based on the AF parameters estimated in experiment 2 and using G_{max} as in Glasberg and Moore (2000). Using that K as an average across all conditions and listeners, we then estimated the only free parameter G_{max} which best predicted the MMD while holding all other parameters constant. Finally, using that G_{max} we predicted the masked thresholds across C's.

B. Parameter estimation

1. The decision variable, K

The decision variable K was calculated based on the masked threshold of an "average" listener tested in experiment 1. Thus, thresholds from experiment 1 were averaged across tested listeners and the model was run for all 36 conditions (9 curvatures × 2 precursors × 2 levels). The same masker and target stimuli as in experiment 1 were used, setting the target levels to represent the average thresholds. In the model, G_{max} was 34 dB, corresponding to the normal-hearing I/O function estimated from masking experiments (Glasberg and Moore, 2000). Even though this choice may appear somewhat arbitrary (given that it was not clear to what extent the MOC system was activated in experiment 1), it was a starting point and we will discuss this choice in Sec. IV B 2 b by investigating the influence of G_{max} on the predictions. The resulting K was then averaged across the conditions.

K was calculated separately for each model (one-step and two-step).[4] For the one-step model, i.e., the compression stage consisting of the BM I/O function only, the decision variable was 2.01 and is denoted as K_1. For the two-step model, i.e., with the IHC I/O function included, the decision variable was 1.06 and is denoted as K_2. Note that the difference between K_1 and K_2 represents differences in model characteristics affecting the signal-to-masker ratio at the model output rather than differences in the listener's decision criterion.

2. The maximum gain, G_{max}

G_{max} is a free parameter required to be fit to the experimental data (Glasberg and Moore, 2000). The goal of the procedure described in this section was to find the G_{max} required to best predict the MMDs observed in experiment 1. Stimuli from experiment 1 with C's of −1 and 0.5 were used and two models (one-step and two-step) were tested.

a. One-step model. First, for each of the two C's, the model was run with masker stimulus and target-and-masker stimulus for target levels from 0 to 110 dB in 0.1-dB steps. The target level for which K converged to K_1 represented the predicted target level at the threshold for the specific C. The predicted MMD was then calculated as the difference between the predicted target levels at threshold obtained for the two C's. This procedure was repeated for G_{max}'s in 1-dB steps from 0 to 70 dB, for two masker levels (60 and 90 dB), and two precursor conditions (on- and off-frequency).

The upper panels in Fig. 6 show the predicted MMDs. Generally, the predicted MMDs were greater in the 60-dB level condition than in the 90-dB condition. This is consistent with the general pattern of results from experiment 1 (in Fig. 6 shown as symbols at negative G_{max}). For the 60-dB level condition, the predicted MMDs showed a maximum of 8.7 dB at G_{max} of 53 dB. This is close to the G_{max} of 50 dB showing the largest difference in $S_{3,M}$ between the two C's (see Fig. 5, and Sec. IV A 3). That difference is in the order of 2 dB, and if one interprets that difference as an MMD, it will appear much smaller than the 8.7 dB shown by model. This apparent discrepancy arises because Fig. 5 shows the

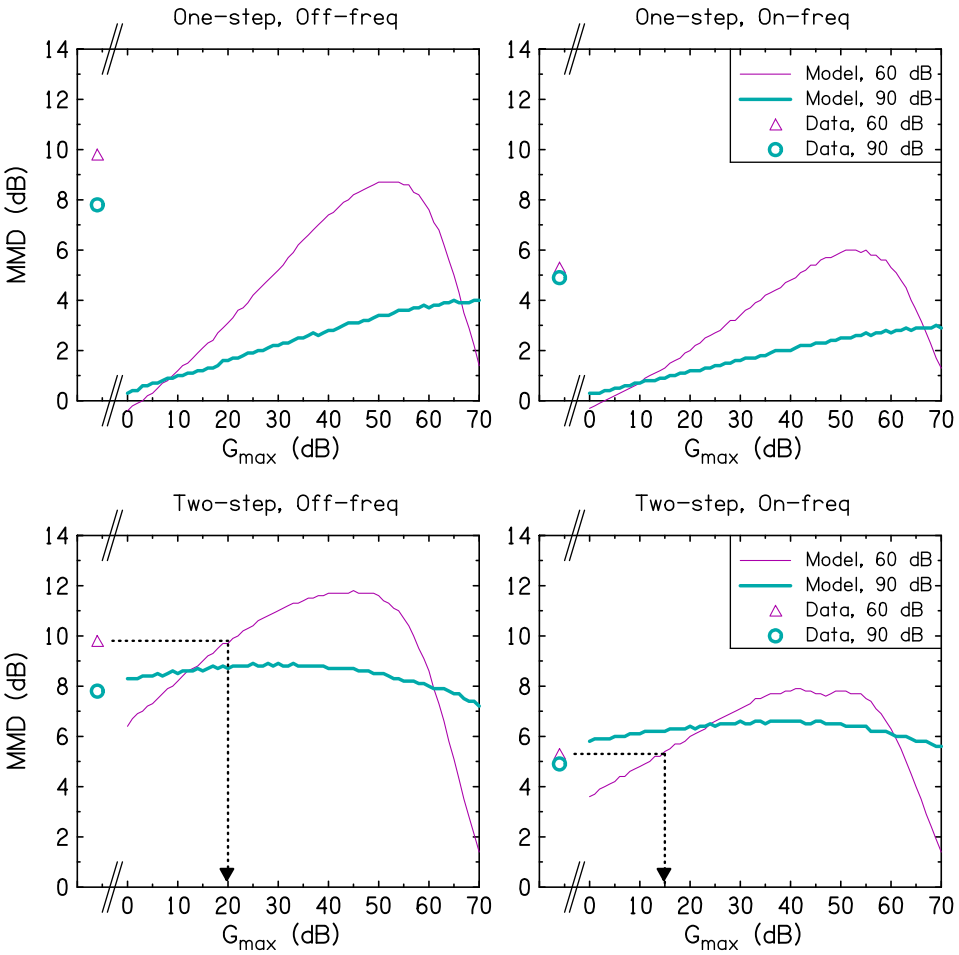

FIG. 6. (Color online) MMD as a function of G_{max} based on the one-step (top row) or two-step (bottom row) model. The lines show predictions in the 60-dB (thin) and 90-dB (thick) conditions when frequency components were weighted by the roex parameters estimated in the *off-frequency* (left column) and *on-frequency* condition (right column) in experiment 2. The symbols show the actual MMDs obtained in experiment 1. The arrows point to G_{max} for which the predicted MMDs correspond to the actual MMDs in the 60-dB condition.

$S_{3,M}$ calculated for the masker only (not considering the target at all), while Fig. 6 shows the predictions considering nonlinear processing of both stimuli, masker and target.

In contrast, for G_{max} larger than 53 dB, the predicted MMDs decrease with increasing G_{max}. For small signal levels, such high G_{max}'s correspond to highly compressive slopes of the BM I/O function [less than ~0.3 dB/dB, see Fig. 5(b)]. Thus, the envelope waveforms might have been overly compressed, resulting in a small difference in $S_{3,M}$ between C's of 0.5 and −1, i.e., peaky and flat envelopes, respectively. This appears reasonable when considering an extremely compressive slope of zero, for which all outputs saturate at a constant level regardless of their envelope shapes.

In most of the conditions, the one-step model underestimated the actual MPE. We presumed that some additional compression may be required to increase the predicted MPE. Given that the underestimation was most pronounced for the 90-dB level conditions, more compression appears to be required particularly at high levels. We accounted for this additional compression in the two-step model.

b. Two-step model. MMDs were obtained for the same conditions and in a similar way as in Sec. IV B 2 a but K_2 was used as the decision variable. The parameters of the IHC I/O function were heuristically selected in order to obtain (1) a larger compression around 90-dB as compared to the one-step model and (2) predictions within the range of the experimental MMDs. The following parameters were found: $\alpha = 100$, $\beta = 1$, $\gamma = 60$, and $Xshift = -20$. With these parameters, the slope of the IHC I/O function decreases from 0.8 to 0.35 dB/dB as $S_3(t)$ increases from 0 to 100 dB. Note that while our selection of the IHC parameters may appear arbitrary, it comprises the general pattern of the IHC I/O function suggesting more compression at high than at low input levels (Lopez-Poveda and Eustaquio-Martín, 2006). Our two-step model should therefore be considered as a feasibility study demonstrating that besides the BM compression, an additional compression source at high levels may help to predict the data as similarly suggested by Carlyon and Datta (1997a). IHC compression as well as BM compression at high input levels (Ruggero *et al.*, 1997) appear to be potential sources.

The predicted MMDs are shown in the lower panels of Fig. 6. While their patterns seem to be similar to those of the one-step model, they are now in the range of the actual (experimentally observed) MMDs. For the 60-dB conditions, the arrows in Fig. 6 point to the optimal maximal gain, G_{opt}, i.e., the G_{max} best predicting the actual MMDs. The G_{opt} for the on-frequency condition is shown to be 5-dB smaller than that for the off-frequency condition.

For the 90-dB conditions, the MMDs were nearly invariant with G_{max}, yielding predictions in the range of the actual MMDs for all tested G_{max}'s. It seems that while the IHC-induced compression increased the predicted MMDs, especially at lower G_{max}'s, the shallow slope of the IHC I/O function at high levels presumably dominated the predictions, and thus weakened the effect of G_{max}.

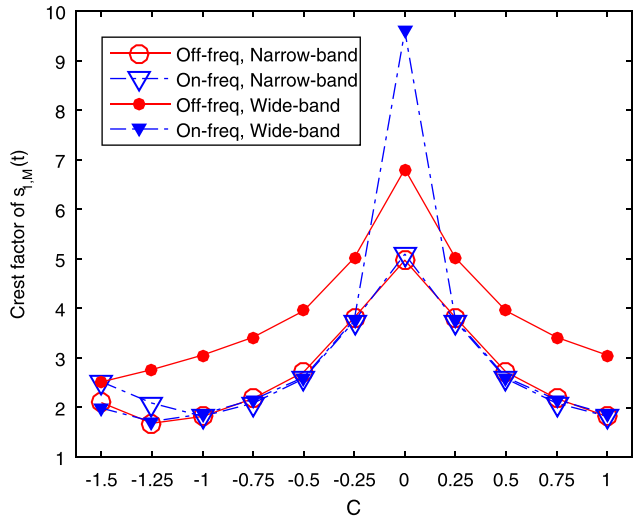

FIG. 7. (Color online) Crest factor as a function of C for a masker waveform after spectral weighting [i.e., calculated based on $s_{1,M}(t)$] using AF parameters estimated for the 60-dB masker SPL in experiment 2. Various conditions for precursors and masker bandwidths are indicated in the inset.

As for the effect of the precursor, the predicted MMDs were generally smaller for the on-frequency condition than for the off-frequency condition. This is in agreement with the actual MMDs. Since the on-frequency precursor broadened the AFs, this is, however, incompatible with the idea that the interaction of a large number of frequency components within the AF yields a peakier internal envelope. To explore this issue, we first calculated the crest factors of masker waveforms filtered by the roex functions from Eq. (3) with parameters estimated from the mean thresholds across listeners for the 60-dB masker SPL in experiment 2. Then, we calculated the crest factors using our maskers and the maskers with a wider frequency range (1.6 to 6.4 kHz, see, e.g., Oxenham and Dau, 2001b). Figure 7 shows the crest factors as a function of C for the two bandwidth and two precursor conditions. For our narrow-band maskers, the differences in crest factor between the two precursor conditions were small (0.4 at most) and consistent with the experimental data. In contrast, for the wide-band maskers, the phase curvature in the on-frequency condition seems to affect the crest factor considerably more than in the off-frequency condition, demonstrating that widening of AFs by the precursor enhances the MPE, as suggested by Oxenham and Dau (2004). A more detailed investigation of those wide-band masker conditions is, however, beyond the scope of this study.

Finally, we re-checked our initial choice of G_{max} (34 dB) used for the estimation of K (see Sec. IV B 1), as it might have influenced the search for G_{opt}. To this end, we reran the parameter estimation for different G_{max}'s, namely, 15, 30, 45, and 60 dB. First, as expected, we obtained different estimates of K_2 (1.4, 1.1, 0.8, and 0.5 dB, respectively). Second, with these K_2's, the predicted MMDs showed similar patterns as functions of G_{max}, even similar to those obtained for $K_2 = 1.06$. The only change was a K_2-dependent shift along the G_{max}-axis. Thus, while with a different initial choice of G_{max} and thus different K_2, we would have obtained a different G_{opt}, the G_{opt} difference between the on-frequency and off-frequency conditions would remain the same. This means that even with a different initial choice of G_{max}, the two-step model would still predict a similar MMD difference between the two precursor conditions, thus reasonably capturing the effect of the precursor on the MPE.

C. Prediction of thresholds for all tested *C*'s

Finally, the two-stage model was used to predict thresholds across all C's for both level and precursor conditions, based on G_{opt}'s from precursor conditions found in Sec. IV. In contrast to the predicted MMDs from Sec. IV, we were interested in the masked thresholds across all C's. Individually for each condition, the predicted thresholds were offset to match the actual thresholds at $C = 0.5$. Importantly, the focus of this modeling approach was to predict the MPE across various conditions, rather than to provide absolute predictions of masked thresholds. In the fitting of the model parameters, most importantly G_{max}, we focused on the MMD and, therefore, any differences in offset across level and precursor conditions can arise from applying those optimum parameters to the entire set of C's. The purpose of normalizing the predictions for each condition is therefore just to facilitate the comparison between data and predictions across C's, as similarly done by Wojtczak and Oxenham (2009). For the 60-dB level conditions, the offsets were 1.06

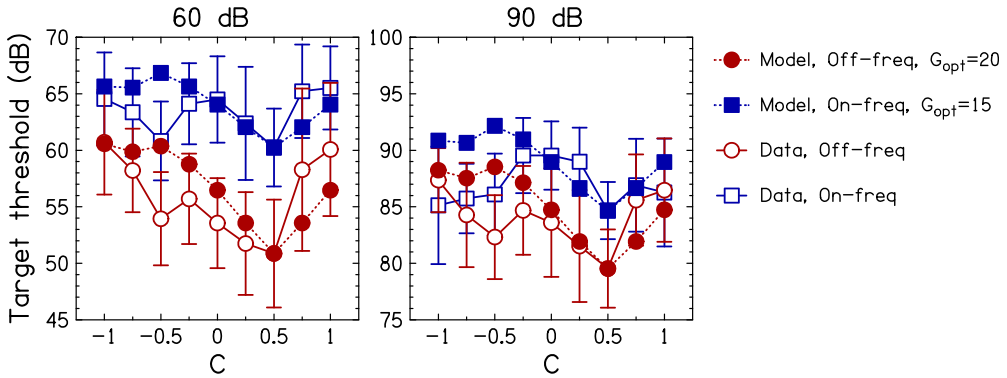

FIG. 8. (Color online) Target thresholds predicted by the two-step model (filled symbols) and obtained from experiment 1 averaged across listeners (open symbols, replication from Fig. 1) across C's for the 60-dB (left panel) and 90-dB conditions (right). The model predictions were offset with respect to the experimental data at $C = 0.5$. The G_{max} values used for the on- and off-frequency precursor conditions (G_{opt} s) are indicated in the legend. Error bars show ±1 standard deviation across listeners.

and 7.64 dB for the off-frequency and on-frequency conditions, respectively. For the 90-dB level conditions, the respective offsets were −2.66 and 0.86 dB.

Figure 8 shows the predicted thresholds and the mean actual thresholds replotted from Fig. 1. The patterns of predicted thresholds well capture the patterns of the actual data. For the 60-dB level conditions, the RMS errors between data and prediction (after the adjustment using the offset parameter) were 3.34 and 2.51 dB for the off-frequency and on-frequency condition, respectively. For the 90-dB level conditions, the respective errors were 2.86 and 3.48 dB. One systematic deviation is, however, that the predicted thresholds are symmetrical with respect to their minimum, whereas the experimental thresholds show shallower slopes towards negative C's as compared to positive C's. The reason for the asymmetry in experimental thresholds is currently not clear.

The model prediction consistently diverged for C around −0.5 where the masked threshold patterns exhibit a local minimum. Interestingly, our recent work on the sensitivity to interaural time differences (ITDs) using similar Schroder-phase harmonic complexes (Tabuchi *et al.*, 2015, 2016) exhibited a locally reduced sensitivity for $C = -0.5$, which was unexpected and could not be predicted by the proposed ITD model. Finding an explanation for these nonmonotonic effects around $C = -0.5$ will be subject of future studies.

V. GENERAL DISCUSSION AND CONCLUSIONS

Experiment 1 examined the influence of a precursor on the MPE at two masker and precursor SPLs (60 and 90 dB). The presence of a precursor was meant to elicit the efferent system, for instance, by means of the MOC reflex, thereby reducing the cochlear gain and compression. This allowed us to test the hypothesis that compression is important in order to explain the MPE. Specifically, it was assumed that an on-frequency precursor activates the MOC system, which linearizes the I/O function for the target following the precursor, and thereby reduces the MPE compared to the condition without a precursor. As a control, a condition with an off-frequency precursor was tested. This condition provided the same temporal stimulus structure as the on-frequency precursor condition, but had most probably no influence on the I/O function for the target and, therefore, no influence on the MPE. The results are overall consistent with our hypothesis, showing a significant reduction of the MPE for the conditions with an on-frequency precursor relative to the conditions with an off-frequency precursor and without a precursor. While nonlinear additivity of masking caused by the on-frequency precursor and the masker might have reduced the MPE as well, an established model of nonlinear masking additivity did not account for the MPE measured in presence of the on-frequency precursor. It therefore appears more likely that the reduction of the MPE due to the precursor is due to a reduction of cochlear compression induced by the efferent system.

Activation of the efferent system also results in a widening of the AF bandwidth (Jennings and Strickland, 2012), which in turn can influence the MPE. Accordingly, the results from experiment 2 showed that the on-frequency precursor expanded the bandwidth of the AF centered on the target frequency. The extent of this bandwidth expansion was at least as great as that found when increasing the masker level from 60 to 90 dB in conditions without the on-frequency precursor. These effects of level and precursor parallel the effects observed for the MPE in experiment 1, suggesting that both the AF bandwidth's widening and the MPE rely on MOC-controlled change of the I/O function. Overall, the results from experiment 2 are in good agreement with previous reports from animal physiology (e.g., Cooper and Guinan, 2006) and human psychoacoustics (e.g., Jennings and Strickland, 2012), which already suggested that OHC gain is important for high frequency selectivity of AFs.

Based on the findings from experiment 1, we proposed a model incorporating peripheral stages of auditory processing, including auditory filtering and compression.[5] Consistent with previous modeling efforts (Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a,b; Summers, 2000; Lentz and Leek, 2001; Shen and Lentz, 2009; Oxenham and Dau, 2001a; Wojtczak and Oxenham, 2009), our model generally underestimated the MPE when accounting only for BM compression. Thus, a two-step model was proposed which combines the BM I/O function with a saturating I/O function. The second compression step was intended to mimic the basic pattern of nonlinear IHC compression. The two-step model appears to better predict the overall patterns of experimental results for both masker levels and precursor conditions, relying on reasonable assumptions about BM and IHC compression.

Our predictions of the MPE generally support the notion that some sort of additional compression is important (Carlyon and Datta, 1997a). In the two-step model, the additional compression was implemented by an I/O function representing the general nonlinear effect of the IHCs, providing more compression at higher than at lower levels. The detailed parameters of the IHC function were chosen heuristically and should thus be studied in more depth in future studies. For example, we assumed the IHC compression to be invariant with efferent activation. Although there is currently no strong evidence of a functional efferent-controlled IHC activity, efferent connections to the IHCs seem to exist (Brown, 2011). Further, other forms of I/O functions, e.g., derived by psychophysical methods such as those based on polynomials (e.g., Plack and Arifianto, 2010), may be considered as the compression stage in our model. Finally, physiological studies reporting compressive BM I/O functions up to rather high levels (Ruggero *et al.*, 1997) should be noted. Assuming such BM I/O functions, IHC compression may not be required at all to predict the MPE effects observed here.

We expected that changes in the AF's bandwidth as a result of modified compression (presentation of on-frequency precursor or increasing masker level) would counteract the effect of the compression itself on the MPE (as suggested by Oxenham and Dau, 2004). In contrast, simulating the effect of spectral filtering in the model showed that widening the AF bandwidth actually slightly *decreased* rather than increased the MPE. Our analysis of crest factors

suggested that, in contrast to our narrow-band maskers, for wider masker bandwidths MPEs would have actually *increased*, consistent with the suggestion by Oxenham and Dau (2004).

Noteworthy, the proposed model currently does not consider other mechanisms potentially involved in the MPE such as combining information across multiple dips according to a multiple-looks mechanism (Viemeister and Wakefield, 1991), suppression effects (e.g., Summers, 2000; Recio and Rhode, 2000; Gockel *et al.*, 2003), or an MOC-induced change of the cochlear phase response (Cooper and Guinan, 2003). While our model predictions suggest that such mechanisms are not required to explain our experimental thresholds, their potential contribution should be studied more directly in future investigations.

ACKNOWLEDGMENTS

We would like to thank the associate editor, Dr. Elizabeth Strickland, and two anonymous reviewers for helpful comments on earlier versions of the manuscript. We thank Michael Mihocic and Jennifer Straub for their assistance in writing software programs for experiments (ExpSuite) and collecting data. This work was supported by the Austrian Science Fund (FWF, Projects Nos. P 24183-N24 and I 1362-N30).

APPENDIX

A well-established model of nonlinear masking additivity (e.g., Plack *et al.*, 2008; Laback *et al.*, 2011) was used to predict masked thresholds for some selected conditions with an on-frequency precursor in experiments 1 and 2. The model assumes that (1) the masked thresholds for the two individual maskers provide a measure of the internal masker excitation, and (2) excitations from the two maskers are added linearly at some higher stage. For equally effective individual maskers (eliciting the same masked thresholds) such a model predicts a 3-dB elevation of masked thresholds for the combined maskers relative to the single maskers if the target is processed linearly. However, if the target is compressed (say by a factor of 2), the masked threshold for the combined maskers has to be elevated by $3 \times 2 = 6$ dB, assuming a constant "internal" masker-to-target excitation ratio at threshold. Such excess masking (beyond linear additivity) occurs because the effect of target compression has to be compensated by a correspondingly larger increase in target level (see Oxenham and Moore, 1995; Plack *et al.*, 2008; Laback *et al.*, 2011). To predict masked thresholds based on masking additivity we first assumed processing of the target by an unadapted BM I/O function, i.e., without MOC activation. To obtain an estimate of the unadapted function, we fitted the BM I/O function in Eq. (5) to the unadapted I/O function reported in Yasin *et al.* (2014) (condition without precursor in their Fig. 4) by varying G_{max}. The resulting "unadapted" G_{max} was 36 dB. Second, we assumed target processing by an adapted I/O function due to the presence of the precursor, i.e., with MOC activation. The adapted I/O function was obtained by fitting Eq. (5) to the I/O function measured by Yasin *et al.* (2014) (their Fig. 4) in the presence of a precursor. Specifically, we considered their 60- and 80-dB SPL conditions (with a 0-ms gap between precursor offset and masker onset) to predict masked thresholds for our 60 and 90 dB masker conditions, respectively, which yielded G_{max}'s of 13 and 23 dB.

In the case of experiment 1, the simultaneous masking effect of the Schroeder-phase masker alone was taken from the no-precursor condition and the forward masking effect of the precursor alone was taken from measurements in experiment 2 (see Sec. III). Figure 9 shows predicted masked thresholds across C assuming full compression of the target, providing a conservative estimate of masking additivity because target compression results in excess masking. The predicted masked thresholds (triangle symbols) show a somewhat flattened pattern across C as compared to the experimental data for the no- or off-frequency precursor conditions, particularly for the 60-dB masker SPL. The predictions are, however, clearly below the experimental data for the on-frequency precursor condition. The diamonds and pentagons in Fig. 9 show predictions assuming reduced target compression due to MOC activation by the precursor,

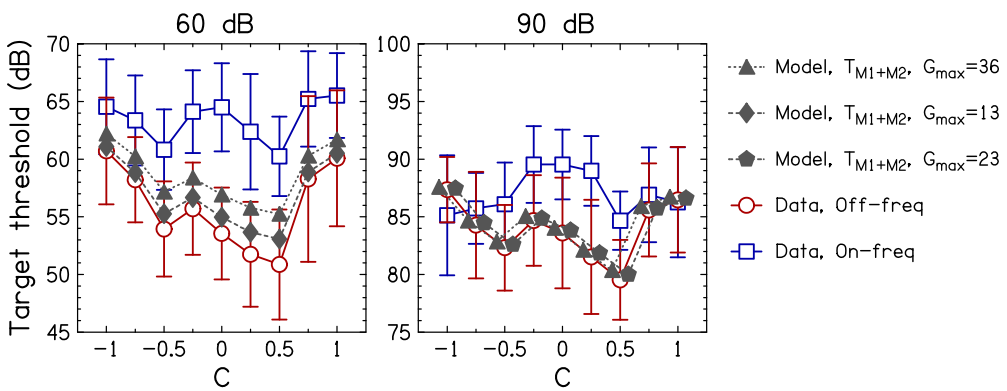

FIG. 9. (Color online) Target thresholds (in dB) predicted by the masking-additivity model (filled symbols connected by dotted lines) and experimental thresholds averaged across all listeners (replicated from Fig. 1, open symbols connected by solid lines) across C's for the 60-dB (left panel) and 90-dB conditions (right panel). T_{M1+M2} denotes the predicted target thresholds in presence of both the precursor (M1) and the Schroeder-phase masker (M2). The meaning of filled symbols is as in Fig. 2. The G_{max} values used for the corresponding conditions are indicated in the legend. Error bars show ±1 standard deviation across listeners. See text for details on the model.

being lower than the predictions assuming the unadapted compression, particularly at the 60-dB SPL. Note that for the conditions of experiment 1 masking additivity might even have been overestimated because some harmonics of the Schroeder-phase maskers fell within the same auditory filter as the target, a condition for which less excess masking is expected [Oxenham and Moore (1995), p. 1932]. Overall, these predictions suggest that while forward masking by the on-frequency precursor causes some flattening of the pattern of masked target thresholds across C, it does not explain the experimentally observed effect of the on-frequency precursor.

The model of masking additivity was also used to predict masked thresholds of experiment 2, considering the precursor and notched noise as maskers. To avoid complications in estimating masking additivity in case of near spectral overlap between notched noise maskers and target (see Oxenham and Moore, 1995), we focused on the normalized notch width of 0.8. Figure 2 shows predictions of the additivity model assuming either full compression of the target (triangles) or reduced compression due to precursor-induced MOC activation (diamonds and pentagons). The predicted masked thresholds were always lower than the measured masked thresholds, suggesting that the elevation of notched-noise thresholds by the on-frequency precursor was not just due to the additivity of masking.

Note that the estimates of compression used for these predictions are based on psychophysical measurements in humans. Physiological data sometimes showed more compressive I/O functions which would result in more excess masking and, thus, potentially better fits to the additivity model.

[1]It should be considered that the remainder of 2π radians (wrapped phase) for each frequency component determines the shape of a Schroeder-phase harmonic waveform, whereas the number of vector rotations (i.e., the multiple of 2π radians) does not affect the waveform shape. The issue of unwrapped phase was also considered in a previous study (Summers *et al.*, 2003).

[2]Minimum and maximum masking conditions were simulated by using C values of -1.5 and 0, respectively, in the model. These correspond to adding the stimulus phase curvatures of -1 and 0.5, corresponding to the maximum and minimum masking, respectively, to a fixed phase curvature of -0.5. The latter value corresponds to the presumed phase curvature of the cochlea.

[3]Although a common form of saturating non-linearity is a second-order Boltzmann function, for the sake of simplicity, we used a first-order Boltzmann function (see also Cooper, 1998).

[4]Note that the parameter K_{AF} for the AF was calculated directly from the mean masked thresholds across listeners in experiment 2 and was completely independent from calculating K_1 and K_2 used in the modeling of experiment 1.

[5]Matlab codes of the current model are available at http://amtoolbox.sourceforge.net/.

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

Bacon, S. P., and Healy, E. W. (**2000**). "Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones," J. Acoust. Soc. Am. **107**, 1589–1597.

Bacon, S. P., Repovsch-Duffey, J. L., and Liu, L. (**2002**). "Effects of signal delay on auditory filter shapes derived from psychophysical tuning curves and notched-noise data obtained in simultaneous masking," J. Acoust. Soc. Am. **112**, 227–237.

Brown, M. C. (**2011**). "Anatomy of olivocochlear neurons," in *Auditory and Vestibular Efferents*, edited by D. K. Ryugo, R. R. Fay, and A. N. Popper (Springer, New York), pp. 17–38.

Carlyon, R. P., and Datta, A. J. (**1997a**). "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., and Datta, A. J. (**1997b**). "Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components," J. Acoust. Soc. Am. **101**, 3648–3657.

Cooper, N. P. (**1998**). "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. **509**, 277–288.

Cooper, N. P., and Guinan, J. J. (**2003**). "Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity," J. Physiol. **548**, 307–312.

Cooper, N. P., and Guinan, J. J. (**2006**). "Efferent-mediated control of basilar membrane motion," J. Physiol. **576**, 49–54.

Dallos, P. (**1985**). "Response characteristics of mammalian cochlear hair cells," J. Neurosci. **5**, 1591–1608.

Glasberg, B. R., and Moore, B. C. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

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

Glasberg, B. R., and Moore, B. C. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.

Gockel, H., Moore, B. C. J., Patterson, R. D., and Meddis, R. (**2003**). "Louder sounds can produce less forward masking: Effects of component phase in complex tones," J. Acoust. Soc. Am. **114**, 978–990.

Guinan, J. J. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.

Hant, J. J., Strope, B. P., and Alwan, A. A. (**1998**). "Variable-duration notched-noise experiments in a broadband noise context," J. Acoust. Soc. Am. **104**, 2451–2456.

Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer Verlag, New York), pp. 412–429.

Hegland, E. L., and Strickland, E. A. (**2016**). "Aging effects on behavioural estimates of suppression with short suppressors," Adv. Exp. Med. Biol. **894**, 9–17.

Jennings, S. G., and Strickland, E. A. (**2012**). "Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity," J. Acoust. Soc. Am. **132**, 2483–2496.

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.

Kohlrausch, A., and 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.

Krull, V., and Strickland, E. A. (**2008**). "The effect of a precursor on growth of forward masking," J. Acoust. Soc. Am. **123**, 4352–4357.

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

Leek, M. R., Dent, M. L., and Dooling, R. J. (**2000**). "Masking by harmonic complexes in budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **107**, 1737–1744.

Lentz, J. J., and 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.

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

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

Lutfi, R. A., and Patterson, R. D. (**1984**). "On the growth of masking asymmetry with stimulus intensity," J. Acoust. Soc. Am. **76**, 739–745.

Oxenham, A. J., and Dau, T. (**2001a**). "Reconciling frequency selectivity and phase effects in masking," J. Acoust. Soc. Am. **110**, 1525–1538.

Oxenham, A. J., and Dau, T. (**2001b**). "Towards a measure of auditory-filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.
Oxenham, A. J., and 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.
Oxenham, A. J., and Ewert, S. D. (**2005**). "Estimates of auditory filter phase response at and below characteristic frequency (L)," J. Acoust. Soc. Am. **117**, 1713–1716.
Oxenham, A. J., and Moore, B. C. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Plack, C. J., and Arifianto, D. (**2010**). "On- and off-frequency compression estimated using a new version of the additivity of forward masking technique," J. Acoust. Soc. Am. **128**, 771–786.
Plack, C. J., Oxenham, A. J., Simonson, A. M., O'Hanlon, C. G., Drga, V., and Arifianto, D. (**2008**). "Estimates of compression at low and high frequencies using masking additivity in normal and impaired ears," J. Acoust. Soc. Am. **123**, 4321–4330.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
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.
Rosen, S., and Baker, R. J. (**1994**). "Characterising auditory filter nonlinearity," Hear Res. **73**, 231–243.
Roverud, E., and Strickland, E. A. (**2010**). "The time course of cochlear gain reduction measured using a more efficient psychophysical technique," J. Acoust. Soc. Am. **128**, 1203–1214.
Roverud, E., and Strickland, E. A. (**2014**). "Accounting for nonmonotonic precursor duration effects with gain reduction in the temporal window model," J. Acoust. Soc. Am. **135**, 1321–1334.
Ruggero, M. A., and Rich, N. C. (**1991**). "Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Rupp, A., Sieroka, N., Gutschalk, A., and Dau, T. (**2008**). "Representation of auditory-filter phase characteristics in the cortex of human listeners," J. Neurophysiol. **99**, 1152–1162.
Russell, I. J., and Sellick, P. M. (**1978**). "Intracellular studies of hair cells in the mammalian cochlea," J. Physiol. **284**, 261–290.
Schroeder, M. (**1970**). "Synthesis of low peak-factor signals and binary sequences of low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.
Shen, Y., and Lentz, J. J. (**2009**). "Level dependence in behavioral measurements of auditory-filter phase characteristics," J. Acoust. Soc. Am. **126**, 2501–2510.
Shera, C. A. (**2001**). "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion," J. Acoust. Soc. Am. **109**, 2023–2034.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and 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.
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. (**2001**). "Overshoot effects using Schroeder-phase harmonic maskers in listeners with normal hearing and with hearing impairment," Hear Res. **162**, 1–9.
Summers, V., de Boer, E., and Nuttall, A. L. (**2003**). "Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects," J. Acoust. Soc. Am. **114**, 294–306.
Summers, V., and 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., and Zenke, K. (**2015**). "Measuring the auditory phase response based on interaural time differences," J. Acoust. Soc. Am. **137**, 2229(A).
Tabuchi, H., Laback, B., Necciari, T., Majdak, P., and Zenke, K. (**2016**). "Modeling the cochlear phase response estimated in a binaural task," in *The 39th Association for Research in Otolaryngology (ARO) Meeting*, San Diego, CA.
Unoki, M., Irino, T., Glasberg, B. R., Moore, B. C., and Patterson, R. D. (**2006**). "Comparison of the roex and gammachirp filters as representations of the auditory filter," J. Acoust. Soc. Am. **120**, 1474–1492.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
von Klitzing, R., and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192–2201.
Walsh, E. J., McGee, J., McFadden, S. L., and Liberman, M. C. (**1998**). "Long-term effects of sectioning the olivocochlear bundle in neonatal cats," J. Neurosci. **18**, 3859–3869.
Walsh, K. P., Pasanen, E. G., and McFadden, D. (**2010**). "Properties of a nonlinear version of the stimulus-frequency otoacoustic emission," J. Acoust. Soc. Am. **127**, 955–969.
Wojtczak, M., Beim, J. A., and 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.
Wojtczak, M., and Oxenham, A. J. (**2009**). "On- and off-frequency forward masking by Schroeder-phase complexes," J. Assoc. Res. Otolaryngol. **10**, 595–607.
Yasin, I., Drga, V., and Plack, C. J. (**2014**). "Effect of human auditory efferent feedback on cochlear gain and compression," J. Neurosci. **34**, 15319–15326.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.