Masking Release for Simultaneous AM Maskers Depends on AM Match of Temporally Flanking Stimulus

Bernhard Laback[1], Hisaaki Tabuchi[2], and Armin Kohlrausch[3]

[1] *Acoustics Research Institute, 1040 Wien, E-Mail:Bernhard.Laback@oeaw.ac.at*
[2] *Department of Psychology, 6020 Innsbruck, E-Mail: Hisaaki.Tabuchi@uibk.ac.at*
[3] *Human-Technology Interaction group, 5600MB Eindhoven, E-Mail: A.Kohlrausch@tue.nl*

Introduction

Simultaneous masking refers to the reduced audibility of a target sound in presence of a simultaneously presented masker sound. It is well known that imposing AM on a masker reduces the amount of masking [1]. The target threshold difference between unmodulated and modulated masker conditions is referred to as masking release (MR). This effect has been most elegantly demonstrated in [2], using so-called Schroeder-phase harmonic complexes as maskers. These stimuli allow to vary the envelope peakedness (i.e., modulation depth) while keeping the power spectrum constant. The MR for shallow versus peaked exemplars of such stimuli amounts up to about 25 dB. The most obvious explanation for the MR is so-called dip listening, i.e., the ability to detect the target in the temporal dips of the masker, having a high signal-to-noise ratio (e.g., [2]).

A recent study [3] investigated the role of another factor potentially contributing to the MR, namely fast cochlear compression. The authors hypothesized that presenting a pure-tone precursor immediately before the masker would reduce compression of the masker stimulus by activating efferent feedback to the cochlea [4], resulting in elevated thresholds for peaked maskers (but not for shallow maskers)[1], thus, reducing the MR. In their experiment the masker was kept short enough to avoid activation of efferent feedback for the masker without precursor. Consistent with the hypothesis, adding a precursor resulted in threshold elevation for peaked maskers but not for shallow maskers.

The first part of the current study (Exps. 1 and 2) re-addressed the potential role of compression and its efferent control by varying the temporal configuration of masker and target stimuli. The results appear to be not compatible with the known temporal dynamics of efferent cochlear feedback, questioning the explanation proposed in [3] that changes in cochlear compression caused the precursor-induced reduction of MR. The second part of the current study (Exps. 3-6) explored an alternative explanation, namely, that adding a flanking signal (such as a precursor) with AM not compatible with the masker's AM disturbs the process of dip listening, thus reducing MR.

Experiment 1

This experiment tested the hypothesis that when increasing the duration of a pair of masker (M) and target (T) stimuli, efferent feedback would reduce compression of later portions of M (after the efferent onset delay of about 50 ms, [4]). Thus, according to the compression hypothesis, increasing the duration of M and the T should make it more difficult (or at least not easier) to detect T compared to a short stimulus not eliciting efferent feedback. T was a 4000-Hz pure tone and M was a Schroeder-phase complex with an F0 of 100 Hz and components ranging from 1600 to 6400 Hz. One independent variable was the peakedness (or modulation depth) of M, determined by the parameter C [3] (see Fig. 1). The second independent variable was the stimulus duration, being either Short (M/T: 40/30 ms) or Long (M/T: 320/300 ms). All stimulus elements had 5-ms raised-cosine ramps. The temporal arrangement of the stimuli is shown in Fig. 2. The masker level was 90 dB SPL. The order of test conditions was randomized across listeners. Thresholds were measured using an adaptive 3-AFC procedure (for details, see [3]). A continuous low-pass filtered noise was used to mask cochlear distortion products. Six normal-hearing listeners served as subjects. Other details were as in [3]. Significance of the effects throughout the study was analyzed using repeated-measures ANOVA, using a significance criterion (α) of 0.05.

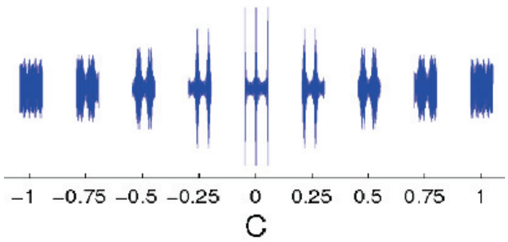

Fig. 1: Excerpts of waveforms of M stimuli used throughout the study with different peakedness (i.e., modulation depth), determined by the factor C. Negative Cs represent time-reversed versions of positive Cs.

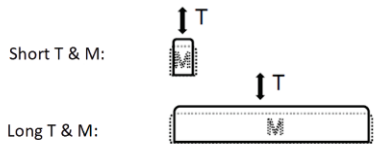

Fig. 2: Temporal arrangement of stimuli of Exp. 1.

Fig. 3 shows masked thresholds as a function of C. The MR was quantified as the difference between maximum (at C=-1) and minimum (at C=0.25) thresholds (indicated in Fig. 3 for condition Long by arrow). Compared to condition Short, condition Long shows significantly *lower* thresholds in case of peaked Ms (but not shallow Ms), resulting in increased MR. This result does not support the idea that reduced compression of a peaked M by efferent activation limits T detection. Rather, the results appear consistent with the idea

[1] *Fast compression is assumed to reduce the excitation level of a modulated sound compared to an unmodulated sound (see [3])*

of multiple-looks like integration [5] of T information across the temporal dips of M.

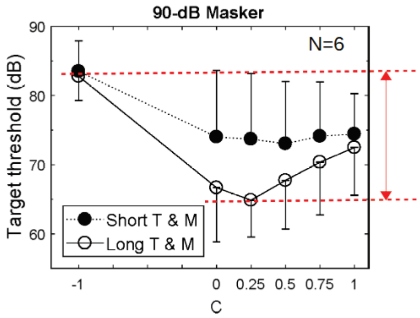

Fig. 3: Results of Exp. 1. Error bars: SDs of the mean. The MR corresponds to the difference of thresholds for conditions C=-1 and 0.25 (as indicated for *Long* by arrow).

Experiment 2

This experiment further tested the potential role of compression and its efferent control by presenting a short T either during the beginning (condition Early) or the end (condition Late) of a long M. According to the efferent-controlled compression hypothesis, MR should be smaller for condition Late. M duration was 320 ms and T duration was 30 ms. The temporal arrangement of stimuli is shown in Fig. 4. Cs of -1 and 0.25 were tested, corresponding to the maximum and minimum threshold, respectively, in Exp. 1. M levels of 60 and 85 dB SPL were tested. Seven normal-hearing listeners were tested. Other details were as in Exp. 1.

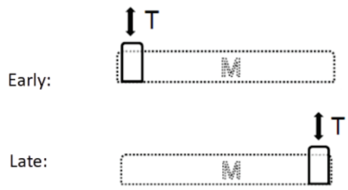

Fig. 4: Temporal arrangement of stimuli of Exp. 2.

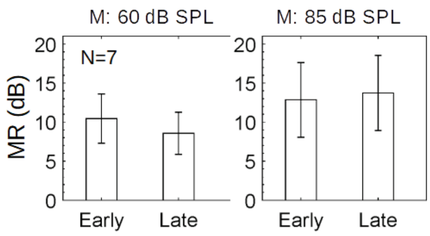

Fig. 5: Results of Exp. 2. MR as defined in Fig. 3. Error bars: SDs of the mean.

Fig. 5 shows no systematic difference in MR between conditions Early and Late, thus not supporting the hypothesis that MR depends on efferent control of compression.

Experiment 3

The results of Exps. 1 and 2 provided no support for the idea that dynamic efferent compression control is responsible for the temporal dynamics of the MR seen with Schroeder-phase maskers. Inspired by some older work on the role of long-term envelope regularity on gap duration discrimination [6], this and the following experiments tested the hypothesis that the MR, considered as the efficiency of dip listening, may depend on some form of long-time AM pattern analysis (denoted here as LAMPA). Specifically, we tested the idea that the efficiency of dip listening depends on the predictability of the masker's AM pattern, as determined by the long-time regularity of the entire stimulus, including any temporally flanking sounds. Assuming that any violation in the regularity of the M's long-time AM pattern would reduce dip listening efficiency, this might explain the reduction in MR when adding an unmodulated PR as in [3]. Given that in [6] is was shown that the envelope analysis window extended both backwards and forwards from the to be judged stimulus part, LAMPA can be expected to be temporally symmetric, thus being equally susceptible to a "disturbing" PR or postcursor (PO).

Exp. 3 tested the effects of adding a PR or PO consisting of either an unmodulated 4-kHz pure tone (Exp. 3a) or a Schroeder-phase complex with either shallow (C=-1) or peaked (C=0.25) AM (Exp. 3b). Bandwidth and level of M (as well as the cursors) were as in Exp. 1. Durations of M, T, and cursors were 40, 30, and 400 ms, respectively. The level of both M and the cursors was 90 dB SPL. The temporal arrangement of stimuli is shown in Fig. 6. Eight normal-hearing listeners were tested. Other details were as in Exp. 2.

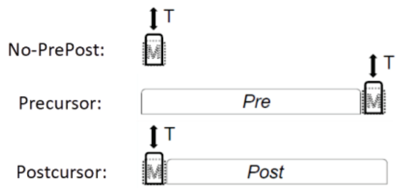

Fig. 6: Temporal arrangement of stimuli of Exp. 3.

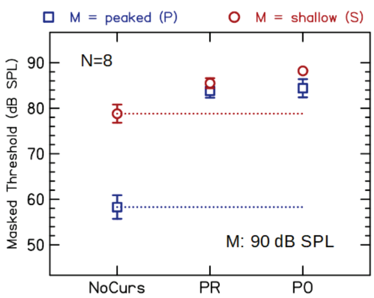

Fig. 7: Results of Exp. 3a, using pure-tone cursors. NoCurs = no cursor, PR=precursor, PO=postcursor. Error bars: SEs of the mean.

Fig. 7 shows that relative to the no-cursor condition, adding a pure-tone PR or PO increased thresholds dramatically for a peaked M (C=0.25) and did much less so for a shallow M (C=-1). MR decreased from about 20 dB without cursor to less than 3 dB with a PR or PO.

Fig. 8 shows the corresponding results for Schroeder-phase maskers. The effect depends critically on the match of the AM depth between M and the cursor (PR or PO); in the unmatched case (Un), masked thresholds considerably increased in case of peaked Ms and did not change in case of shallow Ms. This suggests that adding a cursor with unmatched AM depth specifically disturbs the process of dip listening, but does not generally raise thresholds, as would be expected for an excitation-based mechanism like forward or backward masking. Importantly, these effects are very similar for PRs and POs, consistent with the LAMPA hypothesis.

In the matched case (Ma), even a positive effect (threshold reduction) is observed when adding a PR, but not when adding a PO. This is consistent with a priming-like mechanism. All the reported effects are significant. In summary, the results of Exp. 3 appear consistent with the LAMPA hypothesis.

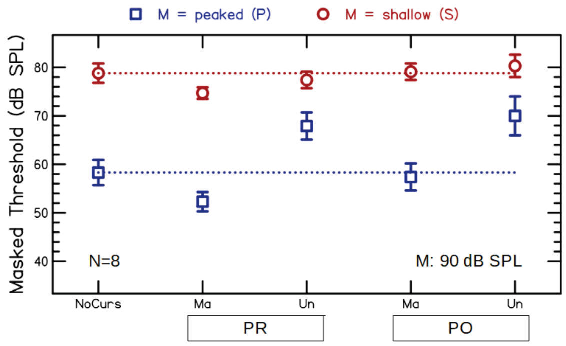

Fig. 8: Results of Exp. 3b using Schroeder-phase pre- and postcursors (PRs and POs). Ma and Un denote matched and unmatched AM depth, respectively, between M and the cursor. Error bars: SEs of the mean.

Experiment 4

The results of Exp. 3 might have been influenced by temporal masking effects, given the lack of a pause between M and the cursors. Further, the very short durations of M and T might limit the generalization of results. To address those issues, Exp. 4 tested the same configurations as in Exp. 3, but using longer durations of M (140 ms) and T (130 ms) and introducing a gap of 200 ms between the PR and M and of 30 ms between M and the PO. These gap durations were chosen as a compromise between avoiding any temporal masking effects and avoiding cognitive effects. Particularly, for a too long gap before the PO, the listener might decide on target detection already before hearing the PO. In addition, pure-tone PR and PO were included (reported in the context of Exp. 6, see Fig. 12). The level of M and the cursors was 80 dB SPL. Eight normal-hearing listeners were tested (five of which were new). All other aspects were as in Exp. 3. Note that the conditions of Exp 4 were tested in balanced order together with those of Exps. 5 and 6, allowing cross-comparisons across these three experiments.

Fig. 9 shows a pattern of results very similar to Exp. 3b (Fig. 8), despite the largely new pool of listeners, the longer durations of M and T, and the inserted gaps between the cursors and M. This suggests a) that the cursor effects do not simply reflect forward or backward masking, b) the robustness of the cursor effects even with a more practically relevant duration of T and M, and c) the generally high reproducibility of the results.

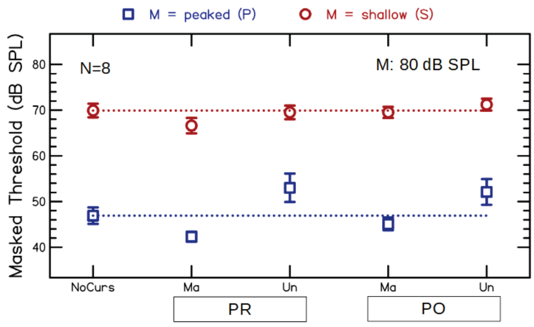

Fig. 9: Results of Exp. 4, introducing a gap between M and the cursors and using longer durations of M and T (see text for details). All other aspects as in Fig. 8.

Experiment 5

This experiment addressed the potential concern with Exps. 3 and 4 that the lack of threshold elevation with unmatched cursor AM for shallow Ms was just a result of the T levels at threshold for the former falling into the less compressive high-level region.[2] Therefore, Exp. 5 tested the PR conditions from Exp. 4 at a 10-dB lower level of M and PR, i.e. at 70 dB SPL (referred to as M70).

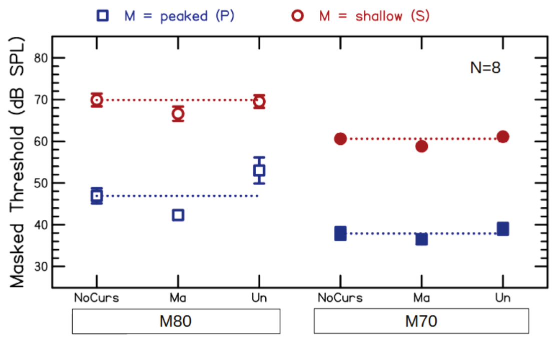

Fig. 10: Results of Exp. 5 (70-dB masker level, M70), compared to PR results of Exp. 4 (80-dB masker level, M80, replicated from Fig. 9). All other aspects as in Fig. 8.

Fig. 10 shows that for M70 (right side), in case of the peaked Ms, the threshold differences between conditions (NoCurs, matched AM depth, unmatched AM depth) are much smaller than for M80 (left side), while the general threshold pattern appears preserved. Given that the peaked M70 thresholds already enter the non-compressive low-level region, these results suggest that the more pronounced threshold differences for M80 are due to cochlear compression. More importantly, in case of the shallow Ms, the threshold differences remain constantly small for both M levels, although M70 thresholds fall into the compressive region. These results support the conclusions of Exp.4 and 5 that the effect of unmatched AM depth of a cursor is specific to peaked Ms, i.e. causes reduced MR.

Experiment 6

The final experiment asked to what extent LAMPA is F0 selective. One question is if any F0 mismatch between M and the cursor *per se* impairs dip listening. Another question is if

[2] *In the compressive mid-level region (with a compression ratio CR), a given change of X dB in masker excitation requires a change of X*CR dB in T level to maintain the threshold criterion.*

the already demonstrated impairment of dip listening by unmatched AM depth is F0-specific, thus occurs only with matched F0. To that end, by varying the F0 of a PR (100, 80, 50 Hz), we varied the F0 congruence of PR with a 100-Hz M. The phase curvature and bandwidth was kept largely constant across F0s by adjusting the constant C (while fulfilling $-1 \leq C \leq 1$). All other aspects were as in Exp. 4.

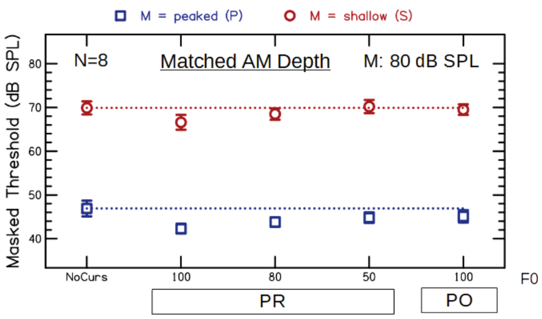

Fig. 11: Effects of adding various cursors with *matched* AM depth, comparing conditions from Exp. 6 (PRs with different F0s denoted) and from Exp. 4 (100-Hz POs, replicated from Fig. 9). All other aspects as in Fig. 8.

Fig. 11 shows the effects of adding precursors with different F0s in case of *matched* AM depth (middle three conditions), including a comparison with PO conditions from Exp. 4 (replicated from Fig. 9). For both peaked and shallow Ms, thresholds *decreased* when adding PR with approximately matched F0 (again suggestive of an overshoot-like effect), but more importantly, no threshold increase for unmatched F0.

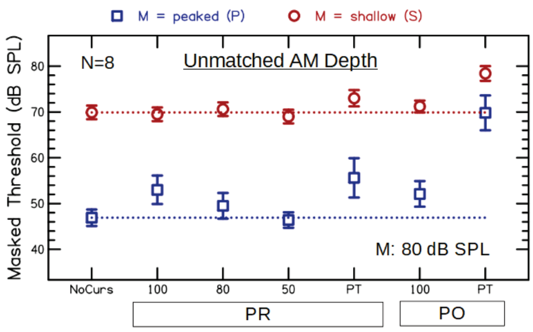

Fig. 12: Effects of adding various cursors with *unmatched* AM depth, comparing various conditions from Exp. 6 (precursors with different F0s denoted) and from Exp. 4 (including pure-tone (PT) cursors not yet reported in Fig. 9). All other aspects as in Fig. 8.

The effects of adding PRs with different F0s in case of *unmatched* AM depth are shown in Fig. 12 (F0s denoted with numbers). Thresholds increased for peaked Ms only when preceded by approximately F0-matched PRs, indicating that the harmful effect of AM depth mismatch is F0 specific. For comparison, also unmodulated pure-tone PRs and POs as well as the 100-Hz POs from Exp. 4 are plotted, all showing significant threshold elevation, albeit differing in magnitude.

Summary and Conclusions

Modulated sounds are well known to elicit much less simultaneous masking than unmodulated sounds, an effect referred to as masking release (MR). Both dip listening and fast cochlear compression have been proposed as mechanisms not mutually exclusively contributing to MR. The first part of the current study (Exps. 1 and 2) re-addressed the potential role of compression and its efferent control by varying the temporal configuration of masker and target. The results were not compatible with the temporal dynamics of efferent compression control. They, thus, question the conclusion of [3] that the reduction of MR as a result of adding a pure-tone precursor was a consequence of compression reduction.

The second part (Exps. 3-6) studied an alternative hypothesis, namely, that any context sound preceding or following the masker, whose envelope shape differs sufficiently from the masker's envelope shape, disturbs the process of dip listening, thus reducing the MR. The observed effects of adding either PRs or POs with matched or unmatched AM depth, were generally consistent with the hypothesis. The addition of unmatched PRs or POs significantly increased masked thresholds for peaked maskers but not for shallow maskers, suggesting AM depth mismatch impairs the process of dip listening. The effect disappeared with increasing F0-mismatch between precursor and masker, tentatively indicating that the effect of AM mismatch is F0 specific. A F0-mismach between M and a cursor *per se* (with matched AM depth) had no effect. The most important finding was that the effects of AM depth mismatch were similar for PRs and POs, even when they were temporally separated from the masker to exclude temporal masking, suggesting the operation of some type of long-time AM pattern analysis (LAMPA) in simultaneous masking by AM sounds. Long-time analysis of the stimulus envelope is exactly the explanation given in [6] for a very different paradigm.

The addition of *matched* precursors was found to *decrease* masked thresholds for both peaked and shallow maskers, reminiscent of the well-known overshoot effect. The lack of such an effect for POs suggests a priming-like mechanism.

The reported effects may have practical implication. The detection threshold of a target sound in presence of a simultaneous masking sound with AM (e.g., voiced speech) would differ by more than 10 dB depending on whether a preceding sound (with a similar F0) had a similar AM depth.

Literature

[1] Buss, S.: Release from masking caused by envelope fluctuations. JASA 78 (1985), 1958-1965

[2] Kohlrausch, A., Sander, A. Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. JASA 97 (1995), 1817-1829

[3] Tabuchi, H, Laback, B., Majdak, P., Necciari, T. The role of compression in the simultaneous masker phase effect, JASA 140 (2016), 2680-2694

[4] Guinan, J. Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans, Ear Hearing 27 (2006), 589-607

[5] Viemeister, N., Wakefield, G. Temporal integration and multiple looks, JASA 90 (1991), 858-865

[6] Münkner, A., Kohlrausch, A., Püschel, D. Influence of fine structure and envelope variability on gap-duration discrimination thresholds, JASA. 99 (1996), 3126-3137