SEQUENTIAL STREAMING REDUCES THE EFFECT OF NON-SIMULTANEOUS MASKING ON AUDITORY INTENSITY RESOLUTION

Patricia Stahn and Daniel Oberfeld
Department of Psychology, Johannes Gutenberg University, 55122 Mainz, Germany
stahn@uni-mainz.de

Abstract

The role of auditory object formation for intensity resolution was investigated by measuring intensity difference limens (DLs) for a pure-tone standard (1 kHz, 30 or 60 dB SPL) under four different masking conditions. All maskers were 935.4-Hz tones with a level of 85 dB SPL. In a two-interval intensity discrimination task, each interval contained a forward masker, a backward masker, or both. Finally, in the streaming condition, the two target tones (standard and standard-plus-increment) were embedded in an isochronous sequence of eight maskers. Maskers and targets in the streaming condition were expected to be perceptually organized as separate auditory streams, which should facilitate selective attention to the targets. In contrast, in the remaining conditions the maskers and the target presented in each interval were expected to be grouped together on the basis of temporal proximity. Compatible with this hypothesis, the masker-induced elevation in the intensity DLs was significantly smaller in the streaming condition than in the remaining conditions. This result supports the hypothesis that the effects of non-simultaneous maskers on intensity discrimination are reduced if the standard and the masker are perceived as separate auditory objects.

Intensity resolution can be strongly affected by nonsimultaneous maskers. These maskers lead to phenomena like the midlevel hump in intensity discrimination: intensity difference limens (DLs) are strongly elevated for a midlevel standard compared to the same situation in quiet, but are hardly elevated for very low and high standard levels (e.g., Zeng, Turner, & Relkin, 1991).

The first attempt to explain the midlevel hump was based on peripheral mechanisms located in the cochlea or in the auditory nerve. It failed on explaining the intensity-difference limens (DLs) elevations caused by contralaterally presented forward maskers, or backward maskers (e.g., Plack & Viemeister, 1992; Schlauch, Clement, Ries, & DiGiovanni, 1999). For this reason, several explanations based on more central processes have been proposed (for a recent overview see Oberfeld, 2008). Based on reports that the similarity between the masker and the standard (e.g. in duration, timbre, pitch, or location) is an important factor in predicting the size of the masker-induced DL elevation (e.g., Oberfeld, 2008; Schlauch, Lanthier, & Neve, 1997), we propose that *attention* might be a useful framework for better understanding the effects of non-simultaneous maskers on intensity resolution.

More specifically, the strong impairment in intensity resolution might be caused by a failure of the listeners to ignore the maskers and selectively attend to the targets. In fact, Oberfeld (2009) showed that listeners use information from to-be-ignored maskers in an intensity discrimination task. In the present experiment, the hypothesis was tested that the perceptual organization of the target and the masker(s) into separate objects (cf. Shinn-Cunningham, 2008) facilitates selective attention to the target. Research on object-based attention in the visual domain (cf. Egeth & Yantis, 1997; Kahneman & Henik, 1981) has demonstrated that it is more difficult to selectively attend to a feature within an object than to attend to one object and ignore another object. Thus, if the masker-target pair presented in a given observation interval is perceived as one object, selectively attending only to the target intensity while ignoring the masker should be more difficult than if the two sounds were processed as two separate auditory objects. To test this idea, we introduced a condition favoring the perceptual organization of the maskers and the targets into two separate auditory streams. We expected the intensity DLs in this *streaming condition* to be lower than the DLs in conditions like the usual forward-masking setting where the maskers and the target presented within a given observation interval can be expected to be grouped together on the basis of temporal proximity. Note that for informational masking, which also represents the effects of central processes, streaming has been reported to strongly reduce the effects of masking (Dau, Ewert, & Oxenham, 2005; Kidd, Mason, & Richards, 2003). We also measured detection thresholds under the different masking conditions.

Method

Participants

Seven students at the Johannes Gutenberg – Universität Mainz participated in the experiment voluntarily (four female, three male; aged 20-28 years). They either received partial course credit or were paid for their participation. All listeners reported normal hearing. For the right ear (the ear tested), detection thresholds measured by Békésy tracking (Békésy, 1947) were better than 20 dB HL between 125 Hz and 16 kHz. Once the topic of the study and potential risks had been explained to them all participants gave written informed consent according to the Declaration of Helsinki. They were uninformed about the experimental hypotheses.

Stimuli and procedure

Intensity discrimination task

Intensity discrimination limens (DLs) were measured using a two-interval, two alternatives forced-choice (2I, 2AFC) adaptive procedure with a 3-down, 1-up rule (Levitt, 1971). The standard was a 1-kHz pure tone with a steady-state duration of 20 ms. The masker was a pure tone presented at a frequency half an equivalent rectangular bandwidth (ERB) below target frequency (935 Hz; Moore & Glasberg, 1983). The steady-state duration of the masker was 50 ms. Both masker and target were gated on and off with 5-ms cosine-squared ramps. The maskers were presented with a longer duration than the targets and at a different frequency in order to promote streaming (see below), and so that the listeners could be instructed simply to judge the short tones (targets) and to ignore the longer tones (maskers). On each trial, there were two observation intervals. An intensity increment was added in-phase to the standard in one of the intervals (selected randomly). Listeners selected the interval containing the louder tone (that is, the standard-plus-increment). Visual trial-by-trial feedback was provided. The sound pressure level of the standard was either 30 or 60 dB SPL.

Intensity DLs were obtained in quiet and in four different masking conditions. The masker level was always 85 dB SPL. In the *streaming condition*, the two target tones (standard and standard-plus-increment) were embedded in an isochronous sequence of maskers presented with an inter-stimulus interval (IOI) of 300 ms (Figure 1, panel A). In this condition, we expected the maskers to be perceived as one stream and the targets as a separate stream. Five maskers were presented before the first target in order to allow for build-up of streaming (e.g., Carlyon, et al., 2010). We expected stream segregation on the basis of the loudness, duration, and pitch differences between maskers and target. The timing of the target onsets in the streaming condition relative to the maskers was so chosen that the masker directly preceding the target was separated from the target by a shorter interval than the masker directly following it, so that forward masking could be assumed to dominate. Nevertheless, the ISI between the target and the "backward" masker directly following the target was sufficiently small for the latter masker to have an effect (Plack, Carlyon, & Viemeister, 1995; Plack & Viemeister, 1992). Therefore, the DLs obtained in the streaming condition were compared

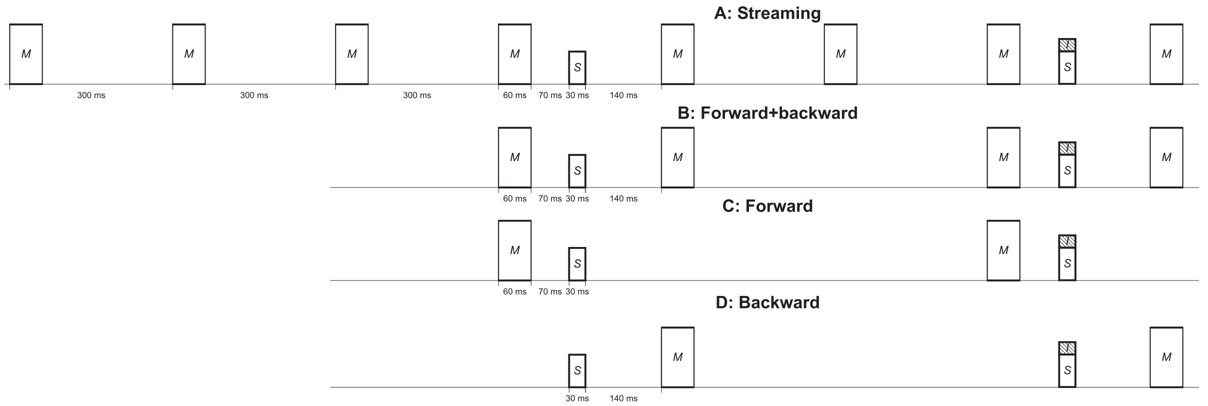

Figure 1. *A*: Streaming condition. The two target tones are embedded in an isochronous masker sequence with an IOI of 300 ms. *B*: Forward-backward: Target with two adjacent maskers. *C*: Forward masker. *D:* Backward masker. Not shown is the in-quiet condition containing no maskers.

with the DLs in a *forward-backward* condition where the targets were presented with their two adjacent maskers (Figure 1, panel B). In this condition, stream segregation was not expected to occur because the "masker rhythm" was interrupted by the relatively long silent interval between the two observation intervals. In the *forward masking* condition (Figure 1, panel C), the silent interval between masker and standard was 70 ms. In the *backward masking* condition (Figure 1, panel D), the silent interval between standard and masker was 140 ms. Listeners were instructed to ignore the maskers. Finally, in the *in quiet* condition no maskers were presented. The stimuli were generated digitally, played back via one channel of an RME ADI/S D/A converter (f_s = 44.1 kHz, 24-bit resolution), attenuated by a TDT PA5 programmable attenuator, buffered by a TDT HB7 headphone buffer, and presented to the right ear via Sennheiser HDA 200 circumaural headphones. The experiment was conducted in a double-walled sound-insulated chamber.

In the adaptive procedure, the initial level of the in-phase intensity increment, expressed in terms of $10\log_{10}(\Delta I/I)$, was 5 dB. The step size was 5 dB until the fourth reversal and 2 dB for the remaining reversals. A track ended when 12 reversals had been obtained or when 80 trials had been presented, whichever occurred first. In each block, two adaptive tracks presenting the same condition were run in a randomly interleaved fashion in order to reduce biases and the predictability of the sequence of target levels (cf. Levitt, 1971). If one of the tracks had already ended (because 12 reversals had occurred) before the termination of the other track, it was still presented with an a priori probability of 0.25. For each of the two tracks, the arithmetic mean of $10\log_{10}(\Delta I/I)$ at all reversals up to the last even-numbered reversal, excluding the first four, was taken as the difference limen corresponding to 79.4% correct. A track was discarded if the standard deviation of $10\log_{10}(\Delta I/I)$ at the counting reversals was greater than 6 dB. For each listener, at least six blocks containing two tracks were obtained for each Standard Level × Masking Condition combination, in separate sessions.

Detection Task

The sound pressure level needed to detect the standard presented in the discrimination task (1 kHz, 30 ms) in quiet and in the four different masking conditions was measured using essentially the same procedure as for the discrimination task. The same masking conditions were presented. In one of the two observation intervals (selected randomly) the signal (1 kHz, 30 ms) was presented. The other interval was empty. The level of the signal was adjusted by a 3-down, 1-up adaptive rule. Listeners select the interval containing the signal and were instructed to ignore the maskers. Visual trial-by-trial feedback was provided. The initial signal level was 30 dB SPL. The step size was 8 dB until the fourth reversal and 2 dB for the remaining eight reversals. The arithmetic mean of the signal levels at the final eight reversals was taken as the detection threshold corresponding to 79.4% correct. For each listener, at least six adaptive tracks were obtained in separate sessions for each masking condition, and at least four tracks in quiet. A track was discarded if the standard deviation of the signal level at the eight final reversals was greater than 6 dB. Time permitting, additional tracks were run if the standard deviation of the DLs estimated in the first six tracks exceeded 5 dB.

Results and Discussion

Detection

Mean detection thresholds in quiet and for the four masking conditions are displayed in Figure 2, Panel A. On average, thresholds were highest in the forward masking condition and in the forward-backward masking condition, followed by the thresholds in backward masking and in the streaming condition. A one-factorial repeated-measure ANOVA using a univariate approach with Huynh-Feldt correction for the degrees of freedom was conducted on the data obtained in the detection task. The Huynh-Feldt correction factor $\tilde{\varepsilon}$ is reported, and partial η^2 is specified as a measure of association strength. The within-subjects factor was masking condition (in quiet, forward, backward, forward-backward, and streaming). There was a significant effect of masking condition, $F(4, 24) = 10.40$, $p = .001$, $\tilde{\varepsilon} = .65$, $\eta^2 = 63$. With the in-quiet condition excluded, the effect of masking condition was also significant, $F(3,18) = 4.00$, $p = .041$, $\tilde{\varepsilon} = .74$, $\eta^2 = .40$. Post-hoc pairwise comparisons between all pairs of masking condition including the in quiet condition were computed using non-pooled error terms (Keselman, 1994) and Hochberg's (1988) sequentially acceptive step-up Bonferroni procedure. For the total of 10 pairs tested, only four tests were significant at an α-level of .05. Thresholds were higher than in quiet for all masking conditions expect backward masking. This pattern is compatible with current models (e.g., Plack & Oxenham, 1998). Surprisingly, however, the detection threshold in the streaming condition was significantly smaller than in the forward masking condition, probably indicating an effect of attention even for the detection task.

Intensity Discrimination

The data were analyzed in terms of the DL-elevation, which denotes the difference between the intensity difference limen ($DL = 10 \log \Delta I/I$) under masking and the DL in quiet. A two-factorial repeated-measures ANOVA with Huynh-Feldt correction for the degrees of freedom was computed. The mean DL-elevation is displayed in Figure 2, panel B, for the four masking conditions and the two standard levels (30 and 60 dB SPL). For all masking conditions except backward masking, the DL-elevation was higher at standard level of 60 dB SPL than at a

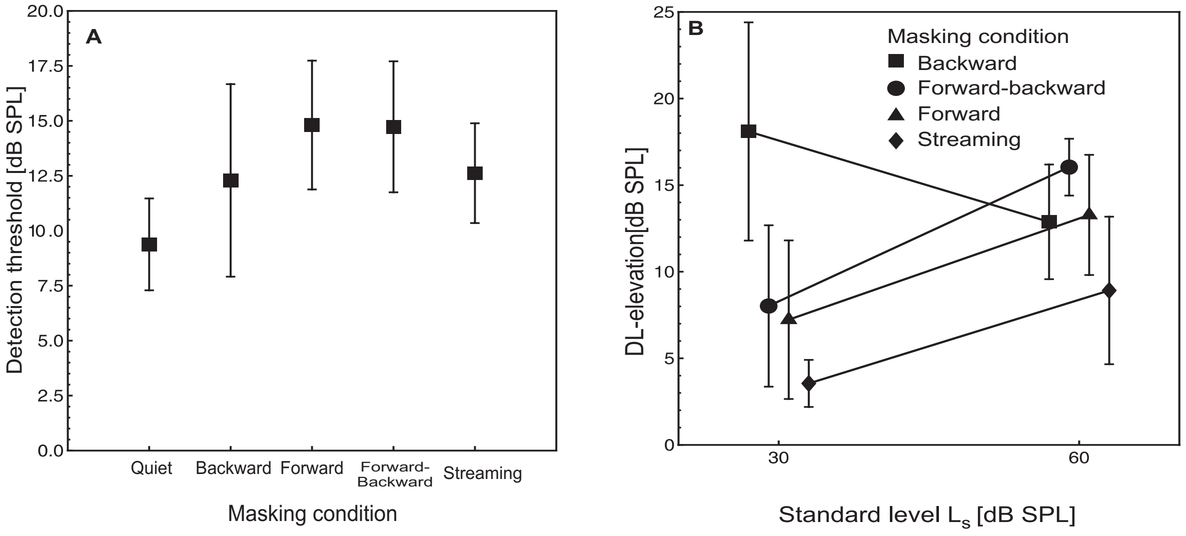

Figure 2. A: Mean detection threshold for the different masking conditions. B: mean DL-Elevation as a function of standard level and masking condition. Error bars show 95% confidence intervals.

standard level of 30 dB SPL, thus demonstrating a midlevel hump (Zeng, et al., 1991). The effect of standard level was significant, $F(1, 6) = 8.91$, $p = .024$, $\eta^2 = .60$. The DL-elevation was minimal in the streaming condition at both standard levels, which is compatible with our hypothesis that the DL-elevation in the streaming condition is reduced because maskers and targets are perceived as two separated streams. This observation was confirmed by a significant effect of masking condition, $F(3, 18) = 24.00$, $p < .001$, $\tilde{\varepsilon} = .77$, $\eta^2 = .80$. The Standard Level × Masking Condition interaction was also significant, $F(3,18) = 5.69$, $p = .007$, $\tilde{\varepsilon} = .99$, $\eta^2 = .49$, most likely owing to the unexpectedly high DL-elevation observed in the backward masking condition at the 30-dB SPL standard level. Note that previous studies also reported extremely high DL elevations for some listeners if a low-level standard was combined with an intense masker (for a discussion see Oberfeld, 2008).We conducted an additional two-factorial ANOVA with the data from the backward masking condition excluded, which again showed a significant effect of masking condition, $F(2, 12) = 21.50$, $p = .001$, $\tilde{\varepsilon} = .76$, $\eta^2 = .78$. Thus, the significant effect of masking condition reported above cannot be attributed merely to the unexpectedly high DL elevation under backward masking at the lower standard level. The Standard Level × Masking Condition interaction with the backward masking condition excluded was now non-significant, $F(2, 12) = 0.36$, $p = .703$, $\tilde{\varepsilon} = 1.00$. Two separate post-hoc ANOVAs comparing the DL elevations in the forward masking condition to the streaming condition and the forward-backward condition, respectively, showed a significant release from masking in the streaming condition ($M = 4.02$ dB, $SD = 3.14$ dB; effect of masking condition $F[1, 6] = 11.44$, $p = .015$, $\eta^2 = .66$), and a marginally significant additivity of masking in the forward-backward condition ($M = 1.78$ dB, $SD = 2.03$ dB; effect of masking condition $F[1, 6] = 5.37$, $p = .060$, $\eta^2 = .47$). Two additional separate post-hoc ANOVAs conducted for the data obtained at the 30-dB SPL and at the 60-dB SPL standard level, again with the backward masking condition excluded, showed a significant effect of masking condition at the 60-dB SPL standard level, $F(2,12) = 9.75$, $p = .003$, $\tilde{\varepsilon} = 1.00$, $\eta^2 = .62$, whereas at the 30-dB SPL standard level the effect was not significant, $F(2, 12) = 2.71$, $p = .107$, $\tilde{\varepsilon} = 1.00$, $\eta^2 = .31$, confirming the observation of a weaker release from masking at the lower standard level (see Figure 2, panel B).

Conclusion

In an intensity discrimination task under non-simultaneous masking, we introduced a condition favoring the perceptual organization of the maskers and the targets into two separate streams. We expected streaming to facilitate selective attention to the targets, compared to a situation where maskers and targets are likely perceived as belonging to one auditory object. The data supported this hypothesis. The masker-induced DL elevation was significantly smaller in the streaming condition than in the remaining masking conditions. The data showed a mid-level hump pattern with higher DL elevations at the 60-dB SPL than at the 30-dB SPL standard level. The release from masking due to streaming was weaker at the lower than at the higher standard level. It remains for future research to show whether the effects of sequential streaming demonstrated here generalize to other conditions promoting the perceptual organization of maskers and targets as separate objects.

Acknowledgements

This work was supported by a grant from Deutsche Forschungsgemeinschaft (DFG) to Daniel Oberfeld (OB 346/4-1: Temporal aspects of auditory intensity processing).

References

Békésy, G. v. (1947). A New Audiometer. *Acta Oto-Laryngologica, 35*(5-6), 411-422.

Carlyon, R. P., Thompson, S. K., Heinrich, A., Pulvermüller, F., Davis, M. H., Shtyrov, Y., et al. (2010). Objective Measures of Auditory Scene Analysis. In E. A. Lopez-Poveda, A. R. Palmer & R. Meddis (Eds.), *Advances in Auditory Research: Physiology, Psychophysics and Models*: Springer.

Dau, T., Ewert, S. D., & Oxenham, A. J. (2005). Effects of concurrent and sequential streaming in comodulation masking release. In D. Pressnitzer, A. de Cheveigné, S. McAdams & L. Collet (Eds.), *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*. New York: Springer.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*, 269-297.

Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika, 75*(4), 800-802.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.

Keselman, H. J. (1994). Stepwise and Simultaneous Multiple Comparison Procedures of Repeated-Measures Means. *Journal of Educational Statistics, 19*(2), 127-162.

Kidd, G., Mason, C. R., & Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *Journal of the Acoustical Society of America, 114*(5), 2835-2845.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2), Suppl 2:467-477.

Moore, B. C., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74*(3), 750-753.

Oberfeld, D. (2008). The mid-difference hump in forward-masked intensity discrimination. *Journal of the Acoustical Society of America, 123*(3), 1571-1581. doi: doi:10.1121/1.2837284

Oberfeld, D. (2009). The decision process in forward-masked intensity discrimination: Evidence from molecular analyses. *Journal of the Acoustical Society of America, 125*(1), 294-303. doi: Doi 10.1121/1.3021296

Plack, C. J., Carlyon, R. P., & Viemeister, N. F. (1995). Intensity discrimination under forward and backward masking: role of referential coding. *Journal of the Acoustical Society of America, 97*(2), 1141-1149.

Plack, C. J., & Oxenham, A. J. (1998). Basilar-membrane nonlinearity and the growth of forward masking. *Journal of the Acoustical Society of America, 103*(3), 1598-1608.

Plack, C. J., & Viemeister, N. F. (1992). Intensity discrimination under backward masking. *Journal of the Acoustical Society of America, 92*(6), 3097-3101.

Schlauch, R. S., Clement, B. R., Ries, D. T., & DiGiovanni, J. J. (1999). Masker laterality and cueing in forward-masked intensity discrimination. *Journal of the Acoustical Society of America, 105*(2 Pt 1), 822-828.

Schlauch, R. S., Lanthier, N., & Neve, J. (1997). Forward-masked intensity discrimination: Duration effects and spectral effects. *Journal of the Acoustical Society of America, 102*(1), 461-467.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182-186. doi: 10.1016/j.tics.2008.02.003

Zeng, F. G., Turner, C. W., & Relkin, E. M. (1991). Recovery from Prior Stimulation II: Effects Upon Intensity Discrimination. *Hearing Research, 55*(2), 223-230.