A review of speech masking release for hearing-impaired listeners with near-normal perception of speech in unmodulated noise maskers

AGNES LÉGER[123], BRIAN C. J. MOORE[4], AND CHRISTIAN LORENZI[123]

[1] *Equipe Audition, Département d'Etudes Cognitives, École normale supérieure, 29 rue d'Ulm, 75005 Paris, France*

[2] *Laboratoire Psychologie de la Perception, Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France*

[3] *UMR CNRS 8158, 45 rue des Saints-Pères, 75006 Paris, France*

[4] *Department of Experimental Psychology, University of Cambridge, Downing Streeet, Cambridge CB2 3EB, UK*

For normal-hearing listeners, intelligibility is higher for speech in fluctuating than in steady noise. This difference is typically reduced for hearing-impaired listeners. It has recently been suggested (Bernstein and Grant, 2009) that the limited benefit for hearing-impaired listeners reported previously results largely from the higher signal-to-noise ratio at which intelligibility was estimated for those listeners in the baseline condition using steady noise. Several studies are reviewed in which normal-hearing and hearing-impaired listeners were tested at identical signal-to-noise ratios; these studies showed limited benefit from noise fluctuations for hearing-impaired listeners, even those with mild losses, despite normal performance in steady noise. Thus, the reduced masking release cannot be explained entirely by the signal-to-noise ratio at which the measurements were made. It also cannot be explained by the amount and configuration of the hearing losses, frequency region, speech material or age.

INTRODUCTION

Normal-hearing (NH) listeners show higher intelligibility for speech in amplitude-modulated noise than in stationary noise, when the overall speech-to-noise ratio (SNR) is held constant. This "masking release" (MR) effect is typically reduced or abolished for hearing-impaired (HI) listeners (Bacon *et al.*, 1998; Eisenberg *et al.*, 1995; Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; 2006a; Lorenzi *et al.*, 2006b; Peters *et al.*, 1998). The reduction has usually been interpreted as indicating that HI listeners have a limited ability to extract information about speech during the dips of fluctuating backgrounds. This may result from reduced audibility and/or supra-threshold deficits in temporal and spectral processing.

Speech intelligibility in the "baseline" stationary noise condition is often poorer than normal for HI listeners, which often led investigators to adjust the SNR to higher (*i.e.*, less adverse) values for HI than for NH listeners. This allowed comparison of

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

the two groups of listeners at identical baseline performance levels in steady noise (Eisenberg *et al.*, 1995; Gustafsson and Arlinger, 1994; 2006a; Lorenzi *et al.*, 2006b). In a recent study, Bernstein and Grant (2009) pointed out that MR is dependent on the baseline SNR (the magnitude of MR decreasing with increasing SNR) and suggested that differences in MR between NH and HI listeners found in previous studies could have been confounded by differences in the SNR at which MR was estimated. They measured MR for NH and HI listeners at identical SNRs rather than at the same performance level, and found that differences in MR between groups were substantially reduced, although the HI listeners still showed some reduction in MR once differences in SNR were controlled for. The residual difference in MR between NH and HI listeners was up to 5 dB. Their results were broadly consistent with a "0-dB rule" suggested by Oxenham and Simonson (2009), according to which substantial MR is only observed when the SNR for speech in steady noise is less than 0 dB.

The results of Bernstein and Grant (2009) illustrate the importance of controlling for SNR differences between groups of listeners before drawing firm conclusions regarding the existence and origin of MR deficits for HI listeners. However, there are some previous studies showing reduced MR for HI listeners despite normal speech reception thresholds (SRTs) and performance level in steady noise maskers. These studies are not subject to the methodological limitation pointed out by Bernstein and Grant (2009) and deserve in-depth re-examination. The present paper reviews these studies and discusses the possible influence of SNR, age, hearing loss, and speech material on the MR deficit.

REVIEW

Table 1 and 2 gives information regarding listener groups, materials, and procedures for each study. For all studies except that of Stuart and Phillips (1996) the noise had the long-term average spectrum of speech. Stuart and Phillips (1996) used white noise, that was lowpass filtered by the earphone. Table 1 gives details of studies reporting performance at fixed SNRs. These studies were selected either because the group data showed no significant difference in identification of speech in steady noise between NH and HI listeners or because some individual HI listeners performed as well as the NH listeners when identifying speech in steady noise (note, however, that there was a trend for better performance for NH than for HI listeners in steady noise in most studies). Results for some of these studies are shown in the left part of Fig. 1; speech identification scores for speech in steady and modulated noise are shown for NH and HI listeners, grouped by SNR. Table 2 gives details of studies in which SRTs were measured, and SRTs for speech in steady noise did not differ significantly for NH and HI listeners. Results for some of these studies are shown in the right part of Fig. 1; SRTs are shown for speech in steady and modulated noise for NH and HI listeners. The columns specify, from left to right: (1) the study; (2) the groups tested and their ages. In all cases except Bacon *et al*. (1998), the groups are those defined by the authors of the studies. For the study of Bacon *et al*. (1998), a sub-group of four HI listeners was selected on the basis that they had SRTs for speech in steady noise similar to those for the NH listeners; (3) the nature of the hearing loss for each group; (4) the type of speech material used (CV = consonant-vowel); (5) the type of masker modulation, modulation rate, modulation depth in percent (*m*), and duty cycle (DC) in cases where square wave modulation was used; (6) the bandwidth of the speech and noise (same for both); (7) the presentation level and the type of amplification (amp.) used for the HI listeners, if any (POGO = prescription of gain and output, approximately a half-gain rule). In the study of Stuart and Phillips (1996) the SRTs in quiet were, on average, 8, 15 and 21 dB SPL for groups NH, HI1, and HI2; (8) the SNRs used.

Study	Age (years, y)	Absolute thresholds *Hearing Loss*	Speech material	Type and depth of masker modulation	Stimulus bandwidth	Presentation level / Amplification for HI	SNR (dB)
Takahashi & Bacon (1992)	NH (n=10): 21-33 y	NH: < 15 dB HL (0.25-8 kHz)	Sentences: SPIN (low context)	Sinusoidal (8 Hz, *m*=100%)	Broadband	Target speech: 70 dB SPL; No amp.	-4 0 +4
	HI1 (n=10): 50-59 y HI2 (n=10): 60-69 y HI3 (n=10): 70-76 y	All HI: 5-15 dB HL (0.25-2 kHz); 40 dB HL (4-8 kHz); *Mild (0.25-8 kHz)*					
Eisenberg *et al.* (1995)	NH (n=8): 19-47 y	NH: < 15 dB HL (0.25-8 kHz)	Sentences: SPIN (high context)	Sinusoidal (31.5 Hz, *m*= 90%)	Broadband	Noise: 76-93 dB SPL after amp. (POGO)	-8
	HI1 (n=8): 37-74 y	All HI: *Mild-mod. Severe (0.25-8 kHz)*					
Jin & Nelson (2006)	NH (n=8): 20-25 y	NH: < 15 dB HL (0.25-8 kHz)	CV & Sentences: IEEE (low context)	Square-wave (8 & 16 Hz, *m*=100%, DC=50%)	Broadband	Target speech: 70 dB SPL for NH; Half-gain amp. for HI	-5 -10
	HI1 (n=9): 20-52 y	HI: 25-55 dB HL (0.5-8 kHz); *Mild-mod. Severe (0.5-8 kHz)*					
Stuart & Phillips (1996)	HI1 (n=12): 55-70 y	HI1: ≤25 dB HL (0.25-4 kHz); >~30 dB HL at 8 kHz	Monosyllabic words NU-6 (low context)	Irregular rectangular (*m*=100%)	Broadband rolling off above 3 kHz	Target speech: 30 dB above SRT in quiet	-10
	HI2 (n=12): 55-70 y	HI2: ≤25 dB HL (0.25-2 kHz); ~50 and 65 dB HL at 4 and 8 kHz; *Mild-mod. Severe (0.25-8 kHz)*					

Table 1: Information about studies in which speech identification performance was measured at one or more fixed SNRs.

Study	Age (years, y)	Absolute thresholds *Hearing Loss*	Speech material	Type and depth of masker modulation	Stimulus bandwidth	Presentation level / Amplification for HI	SNR (dB)
Grose *et al.* (2009)	NH (n=10): 21-29 y	NH: < 5 dB HL (0.25-8 kHz)	Sentences: IEEE (low context)	Square-wave (16 Hz, *m*=100%)	Broadband	Target speech: 65 dB SPL; No amp.	SRT
	HI1 (n=10): 63-75 y	HI: <25 dB HL (0.25-4 kHz); < 40 dB HL (≤ 8 kHz); *Mild (4-8 kHz)*					
Strelcyk & Dau (2009)	NH (n=6): 21-55 y	NH: <20 dB HL (0.25-8 kHz)	Sentences: Dantale II (predictible syntax)	Sinusoidal (8 Hz, *m*=100%)	Lowpass filtered at 1 kHz	Noise: 65 dB SPL; No amp.	SRT
	HI (n=10): 24-74 y	HI: <20 dB HL (0.25-1 kHz); 10-70 dB HL (>1 kHz); *Mild-mod. Severe (1-8 kHz)*					
Bacon *et al.* (1998)	NH (n=11): 22-37 y	NH: < 15 dB HL (0.25-6.3 kHz)	Sentences: HINT (moderate context)	Square-wave (10 Hz, *m*=100%)	Broadband	Noise: 70 dB SPL; No amp.	SRT
	HI (n=4): 38-76 y from 11 HI	HI: <20 dB HL (0.2-1 kHz), 20-70 dB HL (>1 kHz); or 25-60 dB HL (0.2-6.3 kHz); *Mild-moderate (0.2-6.3 kHz)*					

Table 2: Information about studies in which the SRT was measured.

In the figure, circles and squares indicate results for NH and HI listeners, respectively. The bars connecting the symbols are provided for visual guidance, and also to indicate the type of background; filled and open bars indicate steady and amplitude-modulated noise conditions, respectively. Error bars (indicating the overall range of the data) are shown for the study of Jin and Nelson (2006), for which data have been averaged across masker modulation rates (8 and 16 Hz) and starting phases (fixed and random).

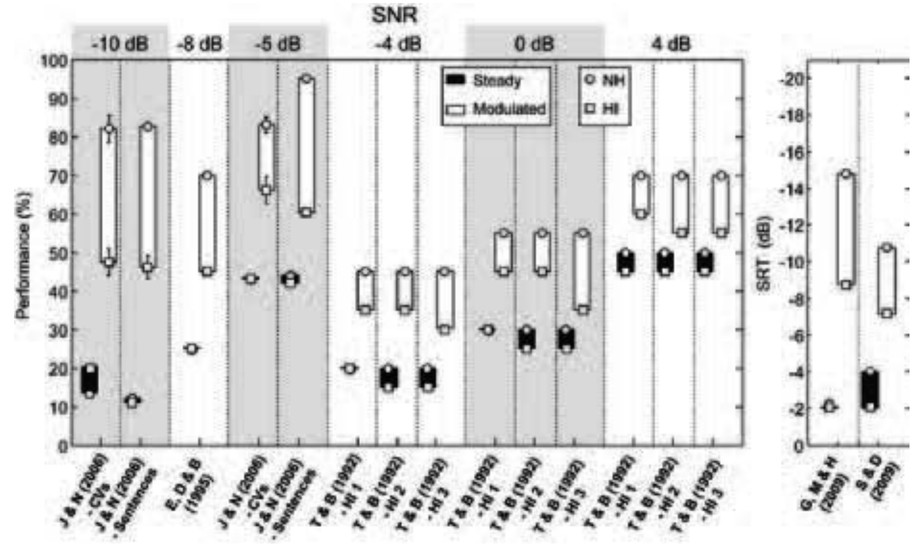

Fig. 1: The left panel shows results from studies in which speech identification performance was measured at one or more fixed SNRs. The SNR is indicated at the top, and results obtained at a given SNR are indicated by a common type of shading (none or grey). The right panel shows results from studies in which the SRT was measured. Results are shown for speech in steady noise (filled bars) and amplitude modulated noise (open bars). Circles and squares show mean scores for NH and HI listeners, respectively. Studies are identified as follows: J & N (2006) – Jin and Nelson (2006); E, D & B (1995) – Eisenberg *et al*. (1995); T & B (1992) – Takahashi and Bacon (1992); G, M & H (2009) – Grose *et al*. (2009); S & D - Strelcyk and Dau (2009). The type of speech material used by Jin and Nelson (2006) was CV (consonant vowel) nonsense syllables or sentences. Speech materials for the other studies are specified in Tables 1 and 2. See text for an explanation of the error bars.

The left part of Fig. 1 shows a general trend for MR to be greater for lower SNRs, as pointed out by Oxenham and Simonson (2009) and Bernstein and Grant (2009). However, for a large range of negative *and* positive SNRs (−10 to +4 dB), HI listeners consistently show less benefit from amplitude fluctuations than NH listeners, despite normal speech identification performance for speech in steady noise. The reduction in benefit is indicated by the length of the open bars. For all of the studies shown in Fig. 1, the difference in MR between the NH and HI listeners was statistically significant and its magnitude was as much as 35% for the SNR of –10 dB. The finding of a MR of about 20 percentage points for the NH listeners and 10% for the HI listeners at an SNR of +4 dB is not consistent with the "0-dB rule" described earlier. The data of Stuart (2008), which are not presented here, are also inconsistent with this rule.

The right part of Fig. 1 again shows that HI listeners benefit less from amplitude fluctuations than NH listeners, despite having normal SRTs in steady noise. In the study of Grose *et al.*(2009), the MR, measured as the difference in SRT between steady and modulated noise, was about 5 dB smaller for the HI than for the NH listeners.

The results shown in Fig. 1 were obtained for groups of 8-10 HI listeners. Inspection of individual data from a study conducted by Bacon *et al*. (1998) (not shown in Fig. 1) indicates that 4 out of their 11 HI listeners (whose characteristics are specified in Table 1) had normal or better-than-normal SRTs for speech in steady noise but nevertheless showed less benefit from amplitude fluctuations (by about 8 dB in terms of SRT) than NH listeners. For two of these listeners, the reduced MR resembled that found for NH listeners when the effect of the hearing loss was simulated using a spectrally shaped steady background noise. However, for the other two HI listeners, the reduction in MR was greater than could be accounted for in this way. This suggests, consistent with the conclusion of Bacon *et al*. (1998), that reduced MR for HI listeners cannot always be accounted for by reduced audibility.

In summary, the results from these studies show that, even when NH and HI listeners show similar performance for speech in steady noise, HI listeners can show less benefit than NH listeners from fluctuations in the noise; both MR measured in percentage points and differences in SRT between steady and modulated noise are smaller for HI than for NH listeners. This is consistent with the idea that HI listeners have either a specific deficit in the ability to "listen in the dips" of background sounds or reduced audibility (despite efforts to made to control for it).

Other demonstrations that SNR is not the only factor determining the amount of MR can be found by comparing results for groups of HI listeners differing in degree of hearing loss. For example, Stuart and Phillips (1996) compared performance in identifying NU-6 words for two groups of HI listeners, with similar ages (both elderly) but differing in their degree of hearing loss above 2 kHz (groups HI1 and HI2 in Table 1). Because of the frequency response of the insert earphone used, the stimuli were effectively lowpass filtered at 3 kHz, so the two groups had similar audiometric thresholds over the frequency range covered by the stimuli. For an SNR of –10 dB, the two groups achieved almost identical scores (15% correct) for speech in steady noise. However, for speech in interrupted noise, the group with the greater high-frequency loss (HI2) achieved 40% correct, whereas group HI1 achieved 50% correct, and this difference was statistically significant. Versfeld and Dreschler (2002) measured SRTs for sentences in steady and fluctuating noise (with a temporal envelope resembling that of a single talker) and included two groups of elderly listeners, one HI and one nearly normal. Most of the listeners in both groups had SRTs of about –4 dB in the steady noise. However, in the fluctuating noise, most listeners in the group with near-normal hearing had SRTs of about –10 dB (*i.e*., they showed a MR of about 6 dB), whereas most listeners in the HI group had SRTs around –4 dB (*i.e*., they showed little or no MR).

The pattern of results described above – reduced MR for HI listeners despite normal performance for speech in steady noise – occurred for a wide range of types of speech materials, including low- and high-predictability sentences, monosyllabic words, and CV nonsense syllables. The effect does not seem to depend critically on frequency region, since it was present in the data of Strelcyk and Dau (2009), obtained using stimuli that were lowpass filtered at 1 kHz. The effect also does not seem to depend critically on age; MR was similar across the three age groups of HI subjects tested by Takahashi and Bacon (1992) and the correlation of MR with age after partialling out the effect of absolute threshold ($r = -0.06$) was not significant. Furthermore, reduced MR for listeners with normal performance for speech in steady noise can be observed for a wide range of amounts and configurations of hearing loss (see column 2 of Table 1). It is noteworthy that reduced MR occurs even for listeners with little or no hearing loss over the frequency range covered by the stimuli.

CONCLUSIONS

There is a general trend for MR to decrease at high SNRs, as described by Bernstein and Grant (2009). However, it seems clear from the review presented here that SNR alone cannot account for the small MR that has often been found for HI listeners. Even when the performance of HI and NH listeners for identifying speech in steady noise is not significantly different, HI listeners, both young and elderly and with varying configurations and degrees of hearing loss, show significantly poorer performance in fluctuating noise, *i.e*., they show reduced MR. Also, clear and significant MR sometimes occurs, for both NH and HI listeners, when the SNR is above 0 dB. Future research should explore the factors other than SNR that lead to reduced MR for HI listeners, such as reduced dip-listening abilities or audibility.

ACKNOWLEDGMENTS

A. Léger, B.C.J. Moore and C. Lorenzi were supported by a grant from the Royal Society (International Joint Project, 2009R3). A. Léger was supported by DualPro Program 7FP-SME222291, and by a CIFRE grant from ANRT and Neurelec. B.C.J. Moore was supported by the MRC (UK). C. Lorenzi was supported by the ANR Presbycusis and RTRS Foundation. We thank Josh Bernstein for helpful comments on an earlier version of this paper.

REFERENCES

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds" J. Speech Lang. Hear. Res., **41**, 549-563.

Bernstein, J. G. W., and Grant, K. W. (**2009**). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners" J. Acoust. Soc. Am., **125**, 3358-3372.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing" J. Speech. Lang. Hear. Res., **38**, 222-233.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing" J. Acoust. Soc. Am., **88**, 1725-1736.
Grose, J. H., Mamo, S. K., and Hall, J. W., 3rd (**2009**). "Age effects in temporal envelope processing: speech unmasking and auditory steady state responses" Ear. Hear., **30**, 568-575.
Gustafsson, H. Å., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise" J. Acoust. Soc. Am., **95**, 518-529.
Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: the effects of temporal resolution" J. Acoust. Soc. Am., **119**, 3097-3108.
Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006a**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure" Proc. Natl. Acad. Sci. USA, **103**, 18866-18869.
Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006b**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception" Int. J. Aud., **45**, 487-495.
Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference" J. Acoust. Soc. Am., **125**, 457-468.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people" J. Acoust. Soc. Am., **103**, 577-587.
Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing" J. Acoust. Soc. Am., **125**, 3328-3345.
Stuart, A. (**2008**). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children" J. Am. Ac. Aud., **19**, 135-146.
Stuart, A., and Phillips, D. P. (**1996**). "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing and presbyacusic listeners" Ear. Hear., **17**, 478-489.
Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly" J. Speech. Lang. Hear. Res., **35**, 1410-1421.
Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners" J. Acoust. Soc. Am., **111**, 401-408.

Prosody perception in simulated cochlear implant listening in modulated and stationary noise

DAVID MORRIS

Institute for Scandinavian Studies and Linguistics, University of Copenhagen, Njalsgade 120, DK-2300, Denmark

Cochlear Implant (CI) listeners can do well when attending to speech in quiet, yet challenging listening situations are more problematic. Previous studies have shown that fluctuations in the noise do not yield better speech recognition scores for CI listeners as they can do for normal hearing (NH) listeners. The aim of this experiment was to investigate the ability of simulated CI listeners in a prosodic task, where F0 Just Noticeable Differences (JND) were measured in modulated and stationary background noise.

A nonsense sentence was created from a recording with durations and overall contour derived from non-scripted Danish speech. The F0 temporal midpoint of the initial syllable was varied stepwise in semitones. Competing signals of modulated white noise and speech shaped noise at 0 dB and 12 dB SNR, were added to the tokens prior to 8-channel noise-excited vocoder processing. Stimuli were presented diotically to 8 NH listeners in a 2AFC task. A question/statement identification experiment was also performed. Results from the JND experiment indicate a significant noise effect for the modulated noise condition at the lower SNR.

INTRODUCTION

It is commonly accepted that CI processing schemes can provide the cues needed for speech recognition in quiet. Other listening situations are more problematic for CI listeners, for instance, speech in competing noise. In a study designed to evaluate the release from masking in fluctuating noise, Nelson et al., (2003) found that CI listeners and simulated CI listeners performed very poorly in a sentence identification task, even at favorable signal-to-noise ratios (SNR). They reported that there was no difference in the identification scores of CI listeners in steady noise and when noise modulation frequencies were between 2-4 Hz, or those approximately corresponding to word and syllable rates. Qin and Oxenham (2003) also found no significant improvement in Speech Reception Threshold (SRT) values between modulated and steady-state noise maskers in 24, 8 and 4 channel CI simulations. These studies highlight the deficit in the ability of CI listeners to 'listen in the gaps' of a temporally modulated masker, an ability that provides release from masking in NH listeners (Festen and Plomp, 1990; Howard-Jones and Rosen, 1993).

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.