

High sentence predictability increases the fluctuating masker benefit

Tim Schoof^{a)} and Stuart Rosen

*UCL Speech, Hearing and Phonetic Sciences, 2 Wakefield Street, London WC1N 1PF,
United Kingdom*

tim.schoof@northwestern.edu, stuart@phon.ucl.ac.uk

Abstract: This study examined the effects of sentence predictability and masker modulation type on the fluctuating masker benefit (FMB), the improvement in speech reception thresholds resulting from fluctuations imposed on a steady-state masker. Square-wave modulations resulted in a larger FMB than sinusoidal ones. FMBs were also larger for high compared to low-predictability sentences, indicating that high sentence predictability increases the benefits from glimpses of the target speech in the dips of the fluctuating masker. In addition, sentence predictability appears to have a greater effect on sentence intelligibility when the masker is fluctuating than when it is steady-state.

© 2015 Acoustical Society of America

[AC]

Date Received: April 17, 2015 Date Accepted: August 12, 2015

1. Introduction

One important factor that allows young normal hearing listeners to understand speech in a considerable amount of background noise is their ability to “listen in the dips” of fluctuating maskers (e.g., Miller and Licklider, 1950). The fluctuating masker benefit (FMB) can be anywhere from a few dB to as much as 20–30 dB depending on the target stimuli used and the temporal characteristics of the masker. The primary aim of the present study was to explore the effects of sentence predictability and masker modulation type on the FMB.

The proportion of the speech signal that is relatively unaffected by the masker is clearly an important factor in successful speech-in-noise perception. The FMB largely depends, for example, on the modulation rate and depth of the masker (Gustafsson and Arlinger, 1994; Wilson and Carhart, 1969). Similarly, noise maskers modulated by the envelope of single or multiple talkers have been shown to result in a smaller release from masking compared to sinusoidally amplitude-modulated (SAM) or square-wave amplitude-modulated (sqAM) maskers (e.g., Bacon *et al.*, 1998).

The FMB likely also depends on the complexity of the target signal and the redundancy of the information available in the amplitude dips of the masker. In other words, the FMB may in part depend on the predictability of the sentences. The glimpses of the target signal in the dips of the masker likely provide contextual cues that reduce the number of options for the target words in the masked portion. Speech perception in noise in general indeed improves when the listener can make use of contextual information (Kalikow *et al.*, 1977). It remains unclear, however, to what extent the FMB depends on contextual information, or sentence predictability.

A secondary aim was to examine whether linguistic closure, the supramodal linguistic ability to fill in missing information, is equally important for the perception of speech in steady-state and fluctuating maskers. Top-down cognitive processing may be particularly important for the perception of speech in fluctuating maskers (Akeroyd, 2008; Rönnberg *et al.*, 2010), because exploiting the information extracted from glimpses of the speech signal may rely more heavily on the ability to fill in missing information. Higher correlations have indeed been found for speech reception thresholds (SRTs) in fluctuating compared to steady-state noise with the Text Reception Threshold (TRT), a measure of linguistic closure, at least for groups of participants extending into middle or older age (see Besser *et al.*, 2013 for a review). However, despite apparent differences in correlation coefficients for the different SRT measures, the correlation coefficients may in fact not have been significantly different from one another. In other words, the ability to fill in missing information may be equally

^{a)}Author to whom correspondence should be addressed. Current address: Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA.

important for the perception of speech in steady-state and fluctuating maskers. Linguistic closure may be important for the perception of speech in steady-state noise because glimpses are to some extent available even in steady-state maskers since speech itself varies in level, causing the short-term signal-to-noise ratio (SNR) to vary.

This study examined the effects of sentence predictability and masker modulation type on the FMB by comparing the FMB for Bamford-Kowal-Bench (BKB) and Institute of Electrical and Electronic Engineers (IEEE) sentences (Rothauser *et al.*, 1969; Bench *et al.*, 1979) in sqAM and SAM noise. First, it was expected that sqAM would lead to a larger FMB than SAM as it leaves a larger proportion of the speech signal unaffected by the masker. Second, the FMB was expected to be larger for the BKB sentences (e.g., “The clown had a funny face”) since they are less complex, contain fewer words, have simpler vocabulary, and are more predictable than the IEEE sentences (e.g., “The birch canoe slid on the smooth planks”). In addition, the importance of the ability to fill in missing information, as measured by the TRT, for the perception of speech in steady-state and amplitude-modulated maskers was assessed.

2. Methods

2.1 Participants

Twenty native British English speakers (age range 18–47 yrs, mean 26 yrs, 11 female) took part. All had normal hearing, defined as pure-tone thresholds of 20 dB HL (hearing level) or better at octave frequencies between 250 and 8000 Hz. None of the participants reported a history of language or neurological disorders. Participants signed a consent form approved by the UCL Research Ethics Committee and were paid for their participation.

2.2 SRT

SRTs were measured for sentences in steady-state (SS) speech-shaped noise that matched the long-term average spectrum of the sentence materials, 10-Hz sqAM speech-shaped noise, and 10-Hz SAM speech-shaped noise (see Rosen *et al.*, 2013, for a description of the speech-shaped noise). The modulation depth was 100% for both modulated maskers. Two different types of sentence materials were used; IEEE and BKB sentences (Rothauser *et al.*, 1969; Bench *et al.*, 1979). The sentences were read by the same male Southern British English speaker. IEEE sentences each contained five key words, while the shorter BKB sentences contained three key words each. The masker always started 500 ms prior to the target sentence and was gated on and off across 100 ms.

Stimuli were presented binaurally at 70 dB SPL (sound pressure level) using an external soundcard (Babyface, RME, Germany). The participants were seated in a soundproof booth and listened to the stimuli over Sennheiser HD 650 headphones. They were asked to repeat what they heard and the experimenter scored correctly repeated key words. No feedback was provided.

The SNR was varied adaptively following the procedure described by Plomp and Mimpen (1979). The first sentence was presented at an SNR of -10 dB. Unless all key words were correctly repeated, the SNR was increased by 6 dB on the next presentation. The initial sentence was repeated until all key words were repeated correctly or the SNR reached 18 dB. The step size was decreased to 4 dB on the following sentence. For each subsequent sentence, the SNR increased by 2 dB when 0 to 2 (IEEE) or 0 to 1 (BKB) key words were correctly repeated or decreased by the same amount for 3 to 5 (IEEE) or 2 to 3 (BKB) correct repetitions, thus tracking 50% of the key words correct in both cases. The number of trials was fixed to 20.

SRTs for each condition were measured twice. A measurement was repeated when fewer than 3 reversals were obtained or when the standard deviation across the reversals exceeded 4 dB at the minimum step size. Thresholds for each run were computed by taking the mean SNR (dB) across the final number of reversals. The SRTs reported here are the mean across the two runs.

Participants were given brief training on sentences in SS, followed by sentences in sqAM, with one of the blocks using IEEE sentences and the other block using BKB sentences. Practice consisted of 3 sentences and started at 0 dB SNR. The order of conditions in the experiment proper was counterbalanced across participants following a Latin square design.

2.3 TRT

The TRT is a visual analogue of the SRT, developed to measure modality general cognitive and linguistic skills associated with speech perception in noise (Zekveld *et al.*,

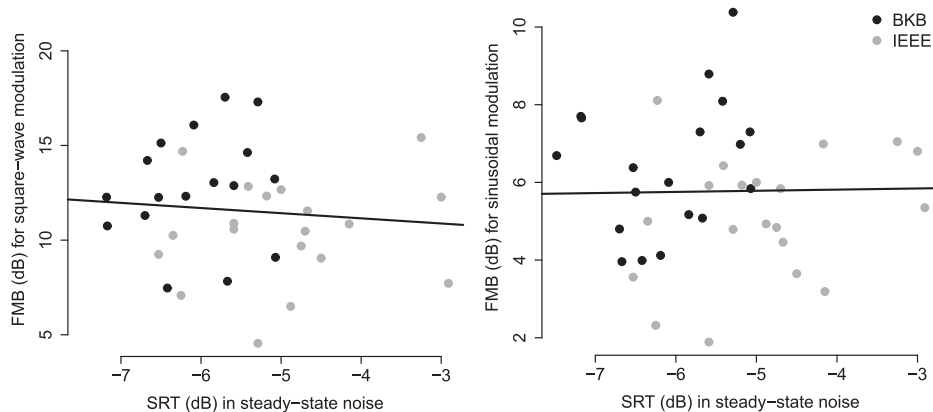


Fig. 1. Scatter plots illustrating the relationship between the FMB and the SRT in steady-state noise for BKB (black dots) and IEEE (light gray dots) sentences, for sqAM (left) and SAM (right) noise. Note the different scales on the y axes.

2007; Besser *et al.*, 2012). Sentences that are partly masked by a vertical bar pattern are presented on a computer screen, with the degree of masking varied from trial to trial by varying the width of the bars.

As in the speech perception in noise task (measuring SRTs), the target stimuli were both IEEE and BKB sentences (Rothauser *et al.*, 1969; Bench *et al.*, 1979). While the target stimuli were taken from the same corpus, the specific sentences used in the two tasks were different. The participants were seated approximately 50 cm from the screen. They were asked to read the sentence out loud. The experimenter scored responses using a graphical user interface which showed all the words in the sentence, as opposed to just the key words as was the case in the SRT task. The scoring screen was not visible to the participants and no feedback was provided.

The degree of masking was varied adaptively following the procedure described by Plomp and Mimpen (1979), tracking 50% correct. The first sentence was presented with 48% unmasked text. Until the sentence was correctly repeated, the percentage of unmasked text was increased by 12% on the next presentation. Subsequent sentences were only presented once. When a sentence was correctly repeated, the degree of masking was increased by 6%. Conversely, the degree of masking was decreased by 6% when a sentence was not repeated correctly.

TRTs were measured in response to 2 lists of 20 sentences each for each sentence type (BKB and IEEE). Data were not included when TRTs across 2 trials differed by more than 8%. The thresholds reported here are the mean of the two trials and reflect percentage of unmasked text (the lower the better).

3. Results

Outliers, defined as data points exceeding the mean by ± 3 standard deviations, were excluded from the analyses. Based on this criterion, three outliers were excluded from the SRT data (two for BKB sqAM, one for IEEE sqAM). Two TRT data points for the IEEE sentences were excluded because the thresholds across the runs differed by more than 8%.

3.1 SRTs

The FMB was calculated by subtracting SRTs in sqAM and SAM noise from SRTs in SS noise for each sentence type separately. It has been suggested, however, that the FMB may partly depend on the SRT in SS noise (Bernstein and Grant, 2009). If this is indeed the case, a potential difference in FMB in terms of sentence type could not necessarily be attributed to differences in the usefulness of glimpses but could in part result from inherent differences in sentence difficulty.

To examine whether FMB across sentence type could be compared directly, two Pearson’s correlations were performed on SRTs in SS and FMB for sqAM and SAM (pooled across the two sets of sentence materials). The results (see Fig. 1) showed that the FMB for sqAM and SAM was not dependent on the SRT in SS noise (sqAM: $r = -0.095$, $p = 0.58$; SAM: $r = 0.018$, $p = 0.92$), which meant a direct comparison of FMB across sentence materials was justified.

Figure 2 illustrates that listeners benefited more from square-wave modulations than sinusoidal modulations in the masker, as indicated by a larger masking release for the sqAM condition. In addition, listeners experienced a larger masking

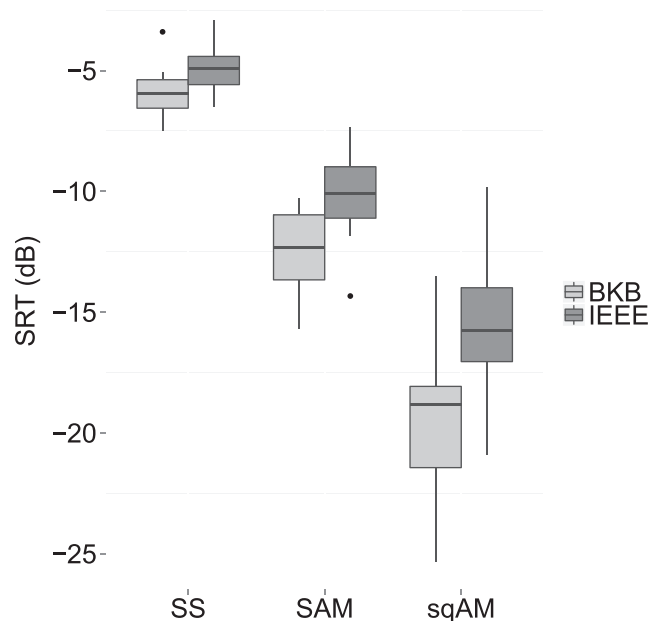


Fig. 2. Boxplots of the SRTs for steady-state (left), sinusoidally (middle), and square-wave (right) amplitude-modulated noise. Results for BKB and IEEE sentences are shown in light and dark gray, respectively.

release for the BKB sentences, compared to the longer and more complex IEEE sentences.

This pattern was confirmed by a mixed effects model on the FMB scores with sentence material and masker modulation type as fixed effects and listener as a random effect. The model showed a main effect of masker type [$F_{(54,1)}=238$, $p < 0.001$, Cohen's $d=2.7$, mean difference = 1.1 dB] and a main effect of sentence type [$F_{(54,1)}=29$, $p < 0.001$, Cohen's $d=1.2$, mean difference = 2 dB]. There was no significant interaction [$F_{(54,1)}=3$, $p=0.07$], which means that the difference in masking release in terms of type of modulation was not more prominent in one sentence type or another.

The results so far have shown that listeners derive a larger FMB for the high-predictability BKB sentences than for the more complex IEEE sentences. The question remains, however, whether sentence predictability also plays an important role in steady-state maskers. To answer this question, a mixed effects model with sentence material and masker type (SS, SAM, sqAM) was performed on the SRTs (not FMBs). The analyses showed a significant interaction between sentence material and masker type [$F_{(2,92)}=6.7$, $p=0.002$]. *Post hoc* paired *t*-tests indicated that SRTs were significantly lower (i.e., better) for BKB compared to IEEE sentences in all maskers (all $p < 0.001$). However, as illustrated in Fig. 2, the effect of sentence type was smallest in steady-state noise (mean difference: SS = 1 dB, SAM = 2.5 dB, sqAM = 3.9 dB; Cohen's d : SS = 1.3, SAM = 1.8, sqAM = 2.1). This suggests that sentence predictability plays a greater role in speech perception in fluctuating than steady-state maskers.

3.2 TRTs

To assess whether TRTs were different across sentence type, a paired *t*-test was conducted. Participants could read the BKB sentences with significantly less available unmasked text compared to the less predictable IEEE sentences [$t_{(17)}=6.3$, $p < 0.001$, Cohen's $d=1.5$; mean percentage unmasked text IEEE 61%, BKB 56%]. These findings are in line with the idea that sentence predictability aids the reconstruction of partially masked speech or text.

3.3 Relationship between TRT and SRT

To answer the question of whether linguistic closure is equally important for the perception of speech in steady-state and fluctuating maskers, two-tailed Pearson's correlations were performed on the SRTs and TRTs. Given the relatively large number of possible correlations that could be performed, simple averages over sentence type were computed for the SRTs in amplitude-modulated (AM) noise, SS noise, and the TRT to reduce the number of measures. The results would otherwise have been subject to very stringent statistical criteria after Bonferroni correction. Note that two participants

were excluded from the correlation analyses because their TRT scores for the IEEE sentences were deemed unreliable and had been removed from the data set. An additional three participants were excluded from the correlation analysis between the TRTs and SRTs in AM noise since their SRT scores for one of the fluctuating noise conditions had been identified as an outlier.

The correlation between the TRTs and SRTs in SS noise was not significant ($r = 0.4$, $p = 0.1$), while that between the TRTs and SRTs in AM noise was ($r = -0.6$, $p = 0.02$). However, this latter relationship was in the wrong direction, suggesting that people with lower (i.e., better) TRT scores have more difficulties understanding speech in a fluctuating masker.

4. Discussion

First, the data showed a larger FMB for BKB than IEEE sentences. Given that the BKB sentences are simpler and contain more contextual information, this finding implies that high sentence predictability leads to a larger FMB. Moreover, the data suggested that sentence predictability plays a more important role in the perception of speech in fluctuating than steady-state maskers. It should be noted, however, that these findings could also be attributed to differences in working memory load across the sentence materials. The IEEE sentences were not only less predictable but also longer and contained more key words than the BKB sentences.

The data furthermore revealed a larger FMB for sqAM than SAM maskers. This can simply be explained by the fact that, at the same modulation rate and depth, sqAM results in a larger proportion of the speech signal being unaffected by the masker. However, the advantage of sqAM over SAM was not more prominent for one sentence type or another. In other words, a relative increase in the size of a glimpse does not aid the use of contextual information.

Last, the results showed that sentence predictability aids the reconstruction of partially masked text. TRT scores were lower (i.e., better) for the simpler BKB compared to the less predictable IEEE sentences. However, this supramodal ability to fill in missing information did not sensibly predict speech perception abilities in either steady-state or amplitude-modulated noise (cf. Besser *et al.*, 2013). While there was a significant correlation between TRTs and SRTs in AM noise, the direction of the relationship was counter-intuitive, with better TRT scores associated with poorer SRTs. The reasons for this unexpected finding remain unclear. The relationship between the TRT and SRT may be more robust, however, in a more heterogeneous population such as older (hearing-impaired) adults.

In sum, the most important finding in this study is that sentence predictability plays a greater role in dip listening than in the perception of speech in a steady-state masker. Furthermore, high sentence predictability improves the benefits of dip listening as well as the ability to read masked text.

Acknowledgments

The authors would like to thank Rebecca Oyekan for her help with data collection, Steve Nevard for technical support, and Adriana Zekveld and J. H. M. Van Beek for providing the TRT test. This work was supported by a Ph.D. studentship grant funded jointly by Action on Hearing Loss and Age UK (Grant No. S19).

References and links

- Akeroyd, M. A. (2008). "Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults," *Int. J. Audiol.* **47**, S53–S71.
- 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.
- Bench, J., Kowal, A., and Bamford, J. (1979). "The Bkb (Bamford-Kowal-Bench) sentence lists for partially-hearing children," *Brit. J. Audiol.* **13**(3), 108–112.
- 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**(5), 3358–3572.
- Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2013). "How linguistic closure and verbal working memory relate to speech recognition in noise—A review," *Trends Amplif.* **17**(2), 75–93.
- Besser, J., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2012). "New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities," *J. Speech Lang. Hear. Res.* **55**, 194–209.
- Gustafsson, H. A., and Arlinger, S. D. (1994). "Masking of speech by amplitude-modulated noise," *J. Acoust. Soc. Am.* **95**(1), 518–529.

- Kalikow, D., Stevens, K., and Elliott, L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**(5), 1337–1351.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**(2), 167–173.
- Plomp, R., and Mimpen, A. (1979). "Improving the reliability of testing the speech reception threshold for sentences," *Audiol.* **18**, 43–52.
- Rönningberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (2010). "When cognition kicks in: Working memory and speech understanding in noise," *Noise Health* **12**(49), 263–269.
- Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (2013). "Listening to speech in a background of other talkers: Effects of talker number and noise vocoding," *J. Acoust. Soc. Am.* **133**(4), 2431–2443.
- Rothausler, E. H., Chapman, N. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Wilson, R. H., and Carhart, R. (1969). "Influence of pulsed masking on the threshold for spondees," *J. Acoust. Soc. Am.* **46**(4), 998–1010.
- Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., and Houtgast, T. (2007). "The development of the text reception threshold test: A visual analogue of the speech reception threshold test," *J. Speech Lang. Hear. Res.* **50**(3), 576–584.