SUPPLEMENTAL MATERIALS

Absorption and enjoyment during listening to acoustically masked stories

Björn Herrmann[1-3*] & Ingrid S. Johnsrude[3,4]

[1]Rotman Research Institute,
Baycrest, M6A 2E1, North York, ON, Canada

[2]Department of Psychology,
University of Toronto, M5S 1A1, Toronto, ON, Canada

[3]Department of Psychology,
University of Western Ontario, N6A 3K7, London, ON, Canada

[4]School of Communication Sciences & Disorders,
University of Western Ontario, N6A 5B7, London, ON, Canada

*Correspondence concerning this article should be addressed to Björn Herrmann, Rotman Research Institute, Baycrest, 3560 Bathurst St, North York, ON, M6A 2E1, Canada. E-mail: herrmann.b@gmail.com

Supplemental analyses for Experiment 2a

For Experiment 2a, participants listened to short snippets (sentences/phrases/words) extracted from a 5-min story from 'The Moth' podcast. The speech snippets were presented under 'clear', +12 dB SNR, or +4 dB SNR conditions (with equal proportions). After each speech snippet, participants reported what they heard by typing it into the computer using the keyboard. The proportion of missed and incorrectly reported words was calculated, here referred to as word-report errors (see also Figure 2 of the main article).

Analyses using non-parametric statistics. Non-parametric Wilcoxon tests comparing median word-report errors between conditions revealed the following effects. Errors were greater for +12 dB SNR than clear ($p = 0.0027$), greater for +4 dB SNR than +12 dB SNR ($p = 1.9\times10^{-5}$), and greater for +4 dB SNR than clear ($p = 6.3\times10^{-6}$) and were higher for non-native compared to native English speakers ($p = 0.0434$).

Analyses using RAU transform. We also used the "rationalized" arcsine transform (RAU; Studebaker, 1985) to transform data closer to a normal distribution, and calculated an ANOVA with the within-subject factor Condition (clear, +12 dB SNR, +4 dB SNR) and the between-subject factor Nativeness (native, non-native English speaker). Word-report errors increased with decreasing signal-to-noise ratio (main effect of Condition: $F_{2,50} = 36.580$, $p < 1\times10^{-6}$; Errors were greater for +12 dB SNR than clear: $t_{26} = 3.094$, $p = 0.005$; greater for +4 dB SNR than +12 dB SNR: $t_{26} = 7.132$, $p < 1\times10^{-6}$; and greater for +4 dB SNR than clear: $t_{26} = 8.802$, $p = p < 1\times10^{-6}$) and were higher for non-native compared to native English speakers (main effect of Nativeness: $F_{1,25} = 10.930$, $p = 0.003$). There was no interaction between Condition and Nativeness ($F_{2,50} = 0.878$, $p = 0.422$).

Analyses limited to data from native English speakers. Non-parametric Wilcoxon tests comparing median word-report errors between conditions revealed more errors for +12 dB SNR than clear ($p = 0.0209$), greater for +4 dB SNR than +12 dB SNR ($p = 8.8\times10^{-5}$), and greater for +4 dB SNR than clear ($p = 8.8\times10^{-5}$)

Supplemental analyses for Experiment 2b

Different groups of people listened to a 5-min story under one of the three masking conditions (clear, +12 dB SNR, +4 dB SNR) and subsequently rated absorption, enjoyment, effort, and comprehension statements. Linear regression models were calculated separately to predict absorption, enjoyment, effort, and comprehension.

Analyses without nuisance predictors. The predictor for each regression was Condition (clear, SNR12, SNR4); in this supplemental analysis, nuisance predictors were not included (see main article for analyses with Nativeness, Sex, and Age included as predictors). Regression analyses did not reveal an effect of Condition on story absorption ($t_{86} = 0.9439$, $p = 0.3479$), suggesting that individuals are similarly

absorbed by an engaging story under clear conditions and moderate masking. Enjoyment was significantly affected by Condition (t_{86} = 2.1347, p = 0.0356): Enjoyment was rated lower for +12 dB SNR (t_{57} = 2.2138, p = 0.0309) and +4 dB SNR (t_{56} = 2.1836, p = 0.0331) compared to the clear condition. The +12 dB SNR and +4 dB SNR conditions did not differ (t_{57} = 0.1327, p = 0.8949). Effort significantly increased as SNR declined (t_{86} = 7.1861, p = 2.3×10^{-10}), such that effort was rated higher for +12 dB SNR (t_{57} = 3.1973, p = 0.0023) and +4dB SNR (t_{57} = 7.5933, p = 3.3×10^{-10}) than for clear, and higher for +4 dB SNR than +12 dB SNR (t_{56} = 3.7470, p = 0.0004). Moreover, story comprehension was affected by Condition (t_{86} = 2.9205, p = 0.0045): comprehension was rated lower for the +4 dB SNR condition compared to clear speech (t_{57} = 2.7686, p = 0.0076).

Analyses for native English speakers. We also examined whether the pattern of results reported in the main article is similar to the results pattern when we limit our analysis to native English speakers. There was no effect of Condition on story absorption (t_{60} = 1.1305, p = 0.2628; none of the other predictors were significant, p > 0.3), suggesting that individuals are similarly absorbed by an engaging story under clear conditions and moderate masking. Enjoyment was significantly affected by Condition (t_{60} = 2.0141, p = 0.0485; none of the other predictors were significant, p > 0.5): Enjoyment was rated lower for +4 dB SNR compared to the clear condition, but this difference was only marginally significant (t_{40} = 1.9584, p = 0.0572). Enjoyment ratings did not differ between +12 dB SNR and clear conditions (t_{39} = 1.6783, p = 0.1013) nor between +12 dB SNR and +4 dB SNR conditions (t_{37} = 0.3623, p = 0.7192). Effort significantly increased as SNR declined (t_{60} = 6.4319, p = 2.3×10^{-8}; none of the other predictors were significant, p > 0.5), such that effort was rated higher for +12 dB SNR (t_{39} = 2.1413, p = 0.0386) and +4dB SNR (t_{40} = 6.6134, p = 6.5×10^{-8}) than for clear, and higher for +4 dB SNR than +12 dB SNR (t_{37} = 3.7646, p = 0.0006). Moreover, story comprehension was affected by Condition (t_{60} = 2.1835, p = 0.0329; none of the other predictors were significant, p > 0.1): comprehension was rated lower for the +4 dB SNR condition compared to clear speech (t_{40} = 2.0790, p = 0.0441). These analyses show that the pattern of results for native English speakers is similar to that reported in the main article (utilizing the combined data from native and non-native English speakers).

Correlations using raw scores. In order to assess whether absorption and enjoyment ratings correlated with each other and/or with effort ratings, Pearson correlations were calculated between absorption ratings and enjoyment ratings and effort ratings, and between enjoyment ratings and effort ratings. To ensure that these analyses are not biased by any mean differences among speech conditions (clear, +12 dB SNR, +4 dB SNR), we subtracted the mean rating from the rating of each individual, separately for each speech condition and measure (absorption, enjoyment, effort) prior to these analyses. We observed a strong correlation between absorption and enjoyment (r = 0.753, p = 2.6×10^{-17}), and moderate negative correlations between absorption and effort (r = –0.293, p = 0.0056) and enjoyment and effort (r = –0.404, p = 9.1×10^{-5}).

Relation between absorption, enjoyment, and effort for native English speakers. In order to investigate for native English speakers whether absorption, enjoyment, and effort share variance that may explain

some of the inter-individual differences, correlations were calculated among these measures (after regressing out Sex and Age). We observed a positive correlation between absorption and enjoyment ($r = 0.755$, $p = 5.75\times10^{-13}$), and negative correlations between absorption and effort (marginally significant; $r = -0.233$, $p = 0.0641$) and enjoyment and effort ($r = -0.329$, $p = 0.008$). These results mirror those reported in the main article where data from native and non-native English speakers were used.

Supplemental analyses for Experiment 3

Participants listened to the audio of a 6-min audiovisual narrative summary of the first seven movies of the Harry Potter franchise in added twelve-talker babble noise at +4 dB SNR. After listening to the summary, participants rated absorption, enjoyment, effort, and comprehension statements. Participants also rated statements about liking Harry Potter and familiarity with Harry Potter from which we calculated a "Harry Potter score".

Correlations without nuisance predictors. Pearson correlations between the Harry Potter score and each of the listening experience measures were calculated (see also Figure 5 of the main article). All four correlations between the Harry Patter score and the listening experience measures were significant: absorption ($r = 0.379$, $p = 0.0055$), enjoyment ($r = 0.459$, $p = 6.1\times10^{-4}$), effort ($r = -0.285$, $p = 0.0406$), and comprehension ($r = 0.657$, $p = 1.2\times10^{-7}$).

Supplemental analyses for Experiment 4

Non-parametric analyses. For each story number (1st, 2nd, 3rd, 4th story) and each measure (motivation, absorption, enjoyment, effort, comprehension), a non-parametric ranksum test was calculated to compare groups (clear, +4 dB SNR). False discovery rate (FDR) was used to account for multiple comparisons (Benjamini and Hochberg, 1995; Genovese et al., 2002). As for the regression analyses in the main article, motivation ($p = 0.0005$), absorption ($p = 0.0002$), and enjoyment ($p = 0.0009$) were lower in the 'noise' group compared to the 'clear' group, but only for the first of four stories (FDR-thresholded). Moreover, listeners in the 'noise group' rated effort higher for the first three stories compared to listeners in the 'clear' group (1st: $p = 8.7\times10^{-6}$; 2nd: $p = 0.0085$; 3rd: $p = 0.0042$; FDR-thresholded). None of the other comparisons were significant.

In order to analyze more directly whether ratings change over time, a linear function was fit to ratings as a function of story number (1st, 2nd, 3rd, and 4th story). Non-parametric ranksum tests revealed more positive slopes for motivation ($p = 0.0023$) and absorption ($p = 0.0171$), and a more negative slope for effort ($p = 0.0034$) for the 'noise' group compared to the 'clear' group.

Analyses limited to native English speakers. The ANOVAs revealed higher motivation, absorption, enjoyment, and comprehension, and lower effort for participants listening to clear stories compared to those listening to stories with added babble (main effect of Group; for all: $F_{1,30} > 6$, $p < 0.02$). The Story Number × Group interaction was significant for motivation, absorption, enjoyment, and effort (for all: $F_{3,90} > 2.8$; $p < 0.05$), but not for comprehension ($F_{3,90} < 2.3$, $p > 0.05$). There were no main effects of Story Number (for all: $F_{3,90} < 2.5$, $p > 0.05$).

Separate regression analyses for each story number revealed that motivation, absorption, enjoyment, and comprehension were lower ($t_{30} < -2.9$, $p < 0.01$) and effort higher ($t_{30} = 5.228$, $p = 1.2\times10^{-5}$) in the 'noise' group compared to the 'clear' group, but only for the first of four stories (FDR-thresholded).

In order to examine whether changes in rating over time differ between groups, a linear function was fit to ratings as a function of story number (1st, 2nd, 3rd, and 4th story), independently for the 'clear' and 'noise' group. The slope was significantly more positive for motivation ($t_{30} = 4.0025$, $p = 0.0004$), absorption ($t_{30} = 3.0567$, $p = 0.0047$), and enjoyment ($t_{30} = 2.332$, $p = 0.0266$), and more negative for effort ($t_{30} = -3.1798$, $p = 0.0034$) for the 'noise' group compared to the 'clear' group, when nuisance variables (Sex and Age) were accounted for.

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57:289-300.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15:870–878.

Studebaker GA (1985) A "rationalized" arcsine transform. Journal of Speech and Hearing Research 28:455-462.