Supplemental Materials

Visualization of the Effect Sizes

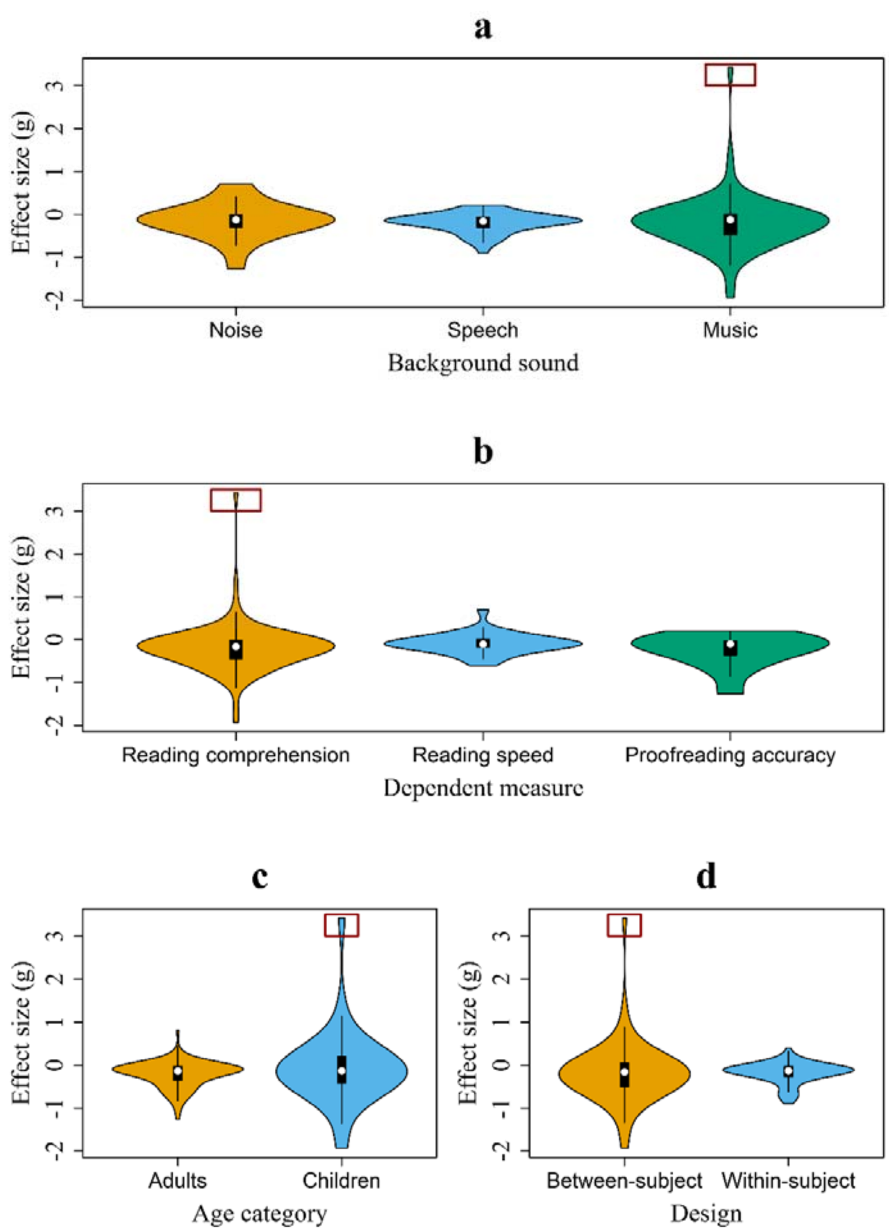

Figure S1. Box plots and probability densities of the effect sizes included in the meta-analysis. Breakdown shown by: background sound type (panel **a**), dependent measure (panel **b**), age of participants (panel **c**), and study design (panel **d**; computed after transforming within-subject effect sizes with Morris & DeShon's, 2002, formula 11). Red rectangle shows one effect size that was excluded as an outlier.

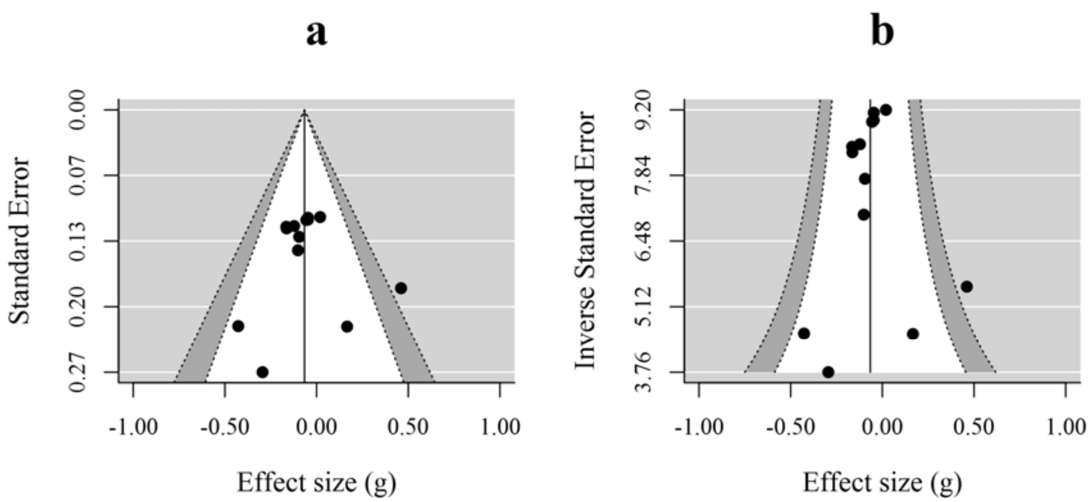

Figure S2. Funnel plot of reading speed effect sizes plotted against their standard error (**a**) and the inverse of their standard error (**b**).

Prior Sensitivity Analysis

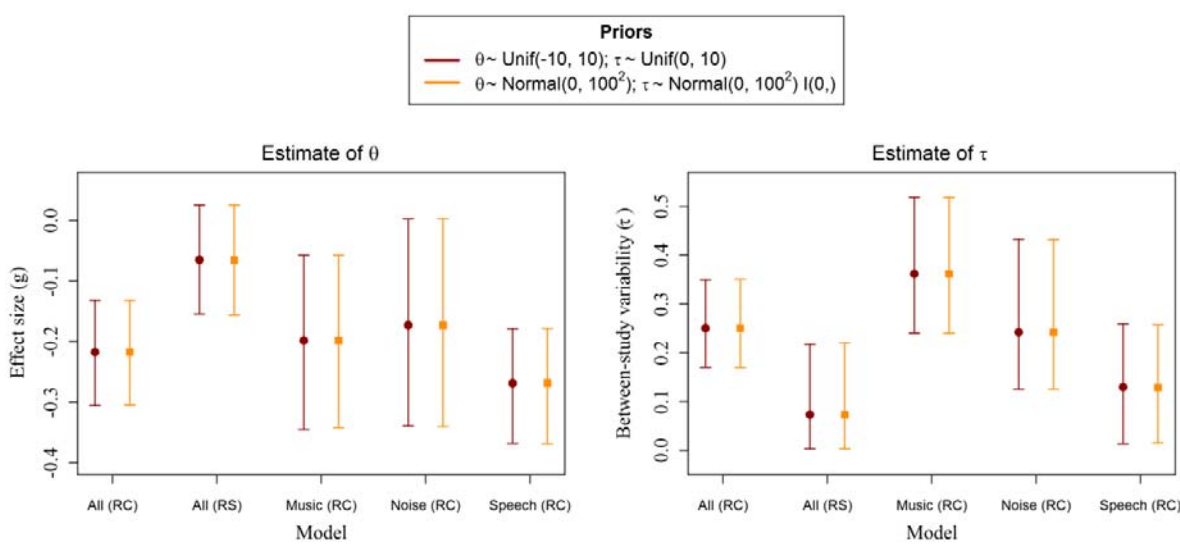

Figure S3. Sensitivity analysis with different priors on the θ and τ parameters for the main meta-analysis results. Uniform priors (dark red) were used in the analysis reported in the main paper. The results show that using diffuse normal priors (orange) did not change the main results reported in the paper. All: all studies. RC: reading comprehension. RS: reading speed. Effective sample size of the MCMC chains for θ (from left to right): 91803, 96976, 20915, 25462, 93678, 100908, 92499, 98585, 47662, 54666. Effective sample size of the MCMC chains for τ (from left to right): 53392, 53451, 18985, 19517, 67441, 68050, 71910, 72202, 11786, 12392.

Robustness Check (Leave-one-out Method)

Robustness analyses were carried out by using the leave-one-out method (see Greenhouse & Iyengar, 2009) to ensure that individual studies did not have very big influence on the effect size estimates. In this method, the meta-analysis is repeated by omitting one different study each time. The summary statistics of the results are reported in Table S1. Overall, the effect sizes changed little by omitting each one of the studies. The effect size range for proofreading accuracy was slightly bigger, but this was likely due to the small number of studies in this analysis (7). This greater variability is not unusual for random-effects meta-analysis with few studies because there is more uncertainty in estimating the between-study variance in the model (cf. Welton et al., 2012).

Table S1

Summary of the Robustness Analysis Using the Leave-one-out Method

Analysis	ES reported in main paper	Leave-one-out			
		Mean ES	SD of ES	Min ES	Max ES
Reading comprehension					
All sounds	-0.21	-0.21	0.006	-0.23	-0.19
Noise	-0.17	-0.17	0.02	-0.19	-0.11
Speech	-0.26	-0.26	0.01	-0.28	-0.24
Music	-0.19	-0.19	0.01	-0.22	-0.16
Reading speed	-0.06	-0.06	0.01	-0.09	-0.05
Proofreading accuracy	-0.14	-0.15	0.04	-0.19	-0.08

Lyrical vs Non-Lyrical Music: Meta-regression Robustness Check

Some of the included studies had effect sizes for both lyrical and non-lyrical music. In order to avoid stochastical dependency among the effect sizes included in this meta-regression analysis, it was necessary to ensure that each study contributed one and only one effect size to either the “lyrical” or “non-lyrical” group. In the paper, the effect sizes were divided into the two groups in a way that maximized the number of effect sizes per group. This is because meta-regressions with larger and more balanced number of observations per group would generally yield more informative results. However, to check for subjectivity in this decision, we did the opposite division of the effect size to compare the results (this will be referred to as the “alternative coding”). The resulting posterior distributions of the mean difference are plotted in Figure S4. As it can be seen, the estimated mean difference was slightly smaller. In the model reported in the paper, there was 95% probability that lyrical music was more distracting than non-lyrical music. For the model with alternative coding, this probability was 83%. Therefore, even though there was slightly more uncertainty and the mean difference was slightly smaller with the alternative coding, our conclusions remain unchanged.

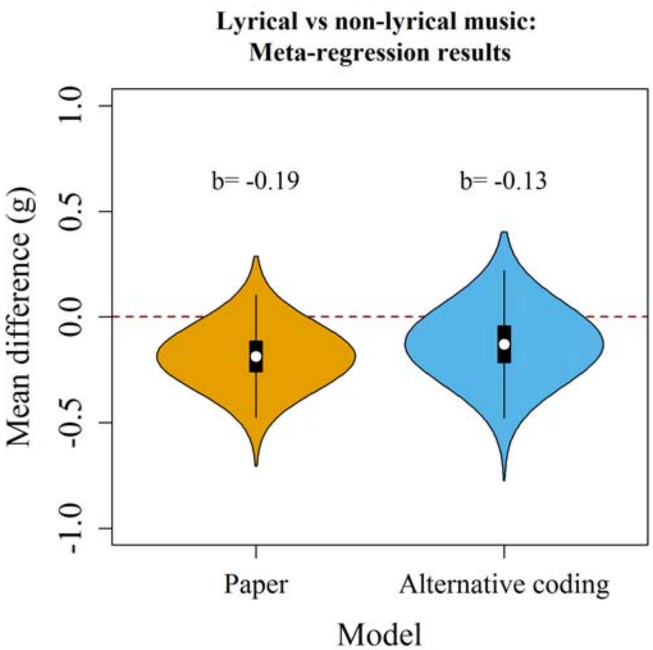

Figure S4. A plot of the posterior distributions of the estimated mean difference in effect sizes between lyrical and non-lyrical music. Plotted are the model reported in the paper (orange) and the model done with the alternative coding of the effect sizes (blue). The results indicate that the decision of which coding to use did not affect the conclusions in the paper. Effective sample size of the MCMC chains for β: 11455 (model reported in the paper), 11695 (model with alternative coding).

Unavailable Data

Due to that fact that four studies did not contain enough information to compute effect sizes and to include them in the meta-analysis, statistical simulations were carried out to explore the consequences of this. The relevant information about these studies is summarised in Table S2. For each study, a realistic interval was computed that should contain the effect size of interest given the available information. The simulations were done by taking 10 000 random draws from a Uniform distribution using the effect size bounds in Table S2. For the variance component, a random draw was also taken from a Uniform distribution with bounds corresponding to the range of variance values in the dataset. The random draws were taken from Uniform distributions to denote ignorance about where on the interval the real value may lie. Each randomly generated effect size was added to the dataset that was analysed in the paper and the meta-analysis was then repeated. The results from the simulations are presented in Table S3 and compared to the effect sizes reported in the main paper. As the simulations show, the results changed very little or not at all when the missing effect sizes were simulated and then added to the analyses. Therefore, the lack of access to the effect sizes of these fours studies did not bias the conclusions from the meta-analysis.

Table S2

Information about Studies with Unavailable Data and Their Anticipated Effect Size

Study	N	Measure	Sound	Available information	Anticipated effect size
Hall (1952)	245	RC	Music	2.37% increase in reading score in the music condition	0< g <0.5
Gawron (1984)	32†	RC	Noise	Effect size known, but not the direction of the difference	|g|= 0.048
Slater (1968)	263	RC	Noise	No sign. differences and "no trends indicative of… [an] effect" (p. 242)	-0.2< g <0.2
Jones et al. (1990), E2	16	PR	Speech	F-value <1; effect size is negative based on the means in Table 2	-0.13< g <0 ‡

RC: reading comprehension accuracy. PR: proofreading accuracy. N: (combined) sample size. All effect sizes are with Morris and DeShon's (2002) correction (where applicable).

† Only two schedules (2x16 participants) are relevant to the analysis

‡ -0.13 is the lowest bound since this would correspond to the effect size when the F-value is 1.

Table S3

Results from the Statistical Simulations with Missing Data (SDs in parenthesis)

Analysis	ES (paper)	Results from 10 000 simulations			
		Mean ES	Range	Mean distribution	Variance distribution
PR	-0.14	-0.13 (0.01)	[-0.15, -0.10]	Uniform(-0.13, 0)	Uniform(0.01, 0.13)
RC: Music	-0.19	-0.19 (0.004)	[-0.20, -0.17]	Uniform(0, 0.5)	Uniform(0.01, 0.20)
RC: Noise	-0.17	-0.16 (0.01)	[-0.17, -0.13]	Uniform(-0.2, 0.2)†	Uniform(0.01, 0.08)

RC: reading comprehension. PR: proofreading.

† Used for Slater's (1968) study. For Gawron's (1984) study, the effect size was positive for half of the simulations (g= 0.048), and negative (g= -0.048) for the remaining half.