

Distraction by unexpected sounds during passage reading and dynamic scene viewing

Martin R. Vasilev^{1*}, Zeynep G. Ozkan², Jinghan Wu³, Julie A. Kirkby⁴, Fabrice B. R. Parmentier^{5,6,7}, Antje Nuthmann⁸

1. Department of Experimental Psychology, University College London, UK
2. Department of Methodology and ERI-Lectura, Universitat de València, Spain
3. Department of Analytics, Technology, and Operations, Leeds University Business School, UK
4. Department of Psychology, Bournemouth University, UK
5. Department of Psychology and Research Institute for Health Sciences (iUNICS), University of the Balearic Islands, Palma, Spain
6. Balearic Islands Health Research Institute (IdISBa), Palma, Spain
7. School of Psychology, University of Western Australia, Perth, Western Australia, Australia
8. Department of Psychology, Kiel University, Germany

* corresponding author at m.vasilev@ucl.ac.uk

ORCID:

Martin R. Vasilev: 0000-0003-1944-8828

Zeynep G. Ozkan: 0000-0002-8684-5789

Jinghan Wu: 0009-0008-0469-9142

Julie A. Kirkby: 0000-0001-6502-0676

Fabrice B. R. Parmentier: 0000-0001-7439-0097

Antje Nuthmann: 0000-0003-3338-3434

Abstract

Previous research suggests that unexpected sounds can inhibit eye-movements, which may occur due to global suppression of motor actions. If this is the case, then eye-movements should be inhibited across a range of tasks, and the degree of inhibition should be relatively independent from the bottom-up perceptual processes associated with each task. In the present study, we tested whether unexpected novel sounds inhibit eye-movements in two complex and ecologically valid tasks: passage reading and dynamic scene viewing. Participants read short passages from textbooks for comprehension and watched short videos of urban environments for memory recall while their eye-movements were being recorded. On every 5th fixation, a sound was played: this was either a repeated “standard” sine wave tone with a probability of .8 or a “novel” environmental sound with a probability of .2. Unexpected novel sounds led to longer fixation durations compared to the standard sound in both tasks, but the effect was slightly weaker in the dynamic scene viewing task. Follow-up analyses suggested that smooth pursuit eye-movements did not fully explain the attenuation of the effect in the scene viewing task. These results suggest that saccadic inhibition by unexpected sounds can occur in different active vision tasks, but the effect may not be entirely independent of the task dynamics.

Keywords: eye-movements, novelty distraction, reading, dynamic scene viewing, videos

Word count: 210 words

Unexpected sensory events elicit brain responses that differ from standard expected events (Näätänen et al., 2010). When we encounter unexpected sounds, the brain processes novel information to determine why it occurs and whether it can be used to inform future decisions (King et al., 2014). Unexpected sounds are thus distracting because they require attention to be diverted away from the current task to process the novel stimuli (Berti, 2008; Escera et al., 1998; Friedman et al., 2001; García-López et al., 2024; Parmentier et al., 2019; Schröger & Wolff, 1998).

Recent research has shown that perceptual surprise can also influence motor control (Wessel, 2017). For example, unexpected sounds have been shown to trigger a global suppression of motor actions (Wessel & Aron, 2013). This suggests that novel sounds may inhibit all types of current motor activities to enhance the processing of the unexpected stimulus. Indeed, the ability to rapidly stop ongoing actions because of perceptual surprise is important in daily life, as it offers flexibility in behaviour adaptation (Dutra et al., 2018).

Wessel and Aron (2017) propose that this motor inhibition occurs through a fronto-basal-ganglia global suppressive network. Such global suppression of motor actions has been shown through Transcranial Magnetic Stimulation, where a loud auditory stimulus (Furubayashi et al., 2000), as well as unexpected novel sounds (Dutra et al., 2018; Wessel & Aron, 2013) resulted in suppressed corticospinal excitability. Moreover, unexpected infrequent sounds have been found to suppress corticospinal excitability even when participants remained still in the absence of any movement initiation (Iacullo et al., 2020). Similarly, unexpected sensory events have also been shown to evoke a tight cortico-muscular coupling (Novembre et al., 2018; Sabu et al., 2024). Therefore, these findings provide evidence of motor suppression at the cortical level and show that sensory detection is reactive rather than just perceptive.

Existing theories and research on the effect of unexpected sounds on motor actions have typically relied on simple binary choice tasks or neurophysiological measures. However, it is still not fully understood how unexpected events affect motor actions in more realistic and ecologically valid tasks. Eye-movements, which are fast and stereotypical motor responses, may provide a good opportunity to test this.

Evidence From Eye-Movements

Recently, eye-tracking has been used to examine the effects of unexpected stimuli on the visual system (e.g., Horstmann & Herwig, 2015; Kadosh & Bonne, 2022; Stephenson et al., 2020). The eyes alternate between fixations, which are used to sample visual information, and saccades, which are used to move gaze from one point of interest to the next (Huey, 1908). Small saccades (known as microsaccades) can also occur within a fixation. These are similar to regular saccades, but much smaller (usually less than 15 minutes of arc; Kowler, 2011). Such small eye movements may help the eyes maintain the acuity of an image during fixations (Martinez-Conde et al., 2009).

Previous research has suggested that oculomotor inhibition can be elicited by external stimuli, as ocular responses are driven by a bottom-up process (Kadosh & Bonne, 2022; Zhao et al., 2019). Unexpected deviant sounds (Kadosh & Bonne, 2022; Widmann et al., 2014) and sometimes even a repeated standard sound (Rolfs et al., 2008) have been shown to inhibit the rate of microsaccades. Additionally, eye movements are inhibited by unexpected sounds when making saccades to simple dot stimuli (Vasilev et al., 2025), while reading single sentences (Vasilev et al., 2019, 2021), during scanning of meaningless sentences (Vasilev et al., 2023), and when viewing digitised paintings (Graupner et al., 2007).

In a picture viewing task, Graupner et al. (2007) presented either auditory or visual distractors on every 5th fixation. The auditory conditions consisted of: 1) 16 “standard” (repeated) sounds and one pitch deviant; 2) 16 standard sounds only; or 3) no distractors. They found that the presentation of the pitch deviant led to an increase in fixation durations. However, the use of a single deviant sound limits the generalisability of their results.

The present study extends the work by Graupner et al. (2007) by using more deviant sounds and by moving beyond static images to dynamic real-world scenes (videos). We also extend existing work on text reading (Vasilev et al., 2019, 2021), which used single sentences as stimuli, by using passages of text. Moreover, we directly compare reading and scene viewing within the same participants.

Dynamic scenes differ from static scenes by including motion and transient visual changes such as flicker, making them more ecologically valid (e.g., Nuthmann & Canas-Bajo, 2022). Roberts et al. (2013) demonstrated that microsaccades, well characterized in controlled fixation tasks, also occur when viewing dynamic naturalistic scenes, extending their relevance beyond simplified viewing conditions. When viewing dynamic scenes, eye-movement behaviour involves not only fixations and saccades but also smooth pursuit movements, which enable continuous tracking of moving objects to maintain stable gaze on them as they shift position over time (Goettker et al., 2020; Goettker & Gegenfurtner, 2021). During free viewing of naturalistic static and dynamic scenes, bottom-up visual features influence gaze allocation; in dynamic scenes, motion and flicker are the strongest independent predictors of gaze location (Mital et al., 2011). Dynamic scenes often elicit a higher degree of “attentional synchrony”, with viewers’ gaze clustering more tightly in space and time compared to static scenes (Dorr et al., 2010; Smith & Mital, 2013). This increased synchrony is thought to reflect stronger exogenous gaze control

during dynamic scene viewing. Finally, dynamic scenes are associated with longer fixation durations compared to static scenes (Dorr et al., 2010; Smith & Mital, 2013).

In the present study, we extend previous work on oculomotor inhibition in static tasks (e.g., reading, picture viewing) to dynamic video scenes. The study had two aims. First, we tested whether Graupner et al.'s (2007) algorithm of presenting sounds on every 5th fixation can be applied more broadly to studying novelty distraction during active vision tasks. Previous studies have mostly used fixed sound presentations (e.g., on specific words or objects on the screen; Rettie et al., 2024; Vasilev et al., 2021, 2023), so it is important to test if this method leads to comparable results. Second, we tested whether unexpected sounds lead to similar inhibition across different tasks. If novel sounds lead to global transient inhibition of motor actions (e.g., Iacullo et al., 2020; Wessel & Aron, 2017), then eye-movements should be inhibited to a similar degree across different tasks. This would be expected if motor inhibition is indiscriminate and not influenced by task dynamics. However, if the degree of inhibition depends on the task, this would suggest that task-specific processes may also influence novelty distraction.

Method

Design

The study had a 2 (*task*: reading comprehension vs dynamic scene viewing) x 2 (*sound*: novel vs standard) design. Each task contained up to 25 items, and the items were blocked by task. The order of the two blocks was counterbalanced across participants. Within each task block, the items were presented in random order. In each task, 50 different novel sounds were presented, randomly dispersed among 200 instances of the standard sound. Thus, the probability of novel sounds was 0.2. Participants performed each task until all sounds were presented, after

which the experiment terminated. This typically meant that not all items were presented for each subject.

Participants

Sixty-four undergraduate students from Bournemouth University took part in the experiment in exchange for course credits (51 female). Their mean age was 21.3 years ($SD= 5.74$ years; range= 18 - 45 years). Participants reported normal (or corrected-to-normal) vision, normal hearing, no prior history of reading or other neurological disorders. The study was approved by the Bournemouth University Research Ethics Committee (ID: 45879).

Power simulations were conducted with the *simr* R package v.1.07 (Green & Macleod, 2016) based on previous data (Vasilev et al., 2021). Assuming an $\alpha= 0.05$ and an effect size of $d= 0.16$, the simulations suggested that about 45-50 participants were sufficient to achieve 95% power of detecting 75% of the expected effect size (see Figure S1 in the Supplemental Materials).

Materials

Reading task. The reading task consisted of 25 short passages taken from undergraduate-level textbooks on various topics (e.g., Biology, Economics, Art, History, Psychology). The passages were 135.1 words long on average ($SD= 7.2$ words; range= 122-152 words) and were followed by a single True/False comprehension question. For example, the questions for the passages in Figure 1a were: “According to the text, a person who smokes marijuana would be considered a bad citizen. True/False” and “According to the text, napping can bring norepinephrine levels back to their normal range. True/False”. The text was formatted in Consolas, a monospaced font, with double line spacing, spanning 10-11 lines. Each letter was 14 pixels wide (0.343°) and printed in black on a white background. The text was left-aligned and

offset by 350px horizontally and 200px vertically on the screen. Participants were instructed to read the passages silently for comprehension.

Scene viewing task. The scene viewing task consisted of 25 short videos of various urban environments (e.g., tourists in front of the Eiffel tower, a tram station, city traffic, a marketplace, an airport). The videos were all in the public domain under a free-to-use license and sourced through the websites <https://pixabay.com> and <https://www.pexels.com>. Their average duration was 30.1 s ($SD= 6.92s$; range= 19 – 48 s). The sound was removed from all videos. Their resolution was 1920 x 1080 pixels (24 fps). Each video was followed by a single Yes/No memory question (e.g., Did any people wait at the bus stop? Did the street have any traffic lights?). Participants were instructed to watch the videos and remember their contents as much as they can.

A) Reading

When trying to avoid a cold or the flu, people are willing to spend a lot of money on immune support supplements and vitamins, but one of the strongest ways to provide powerful immune support is free. After a night of poor sleep, antiviral molecules such as interleukin-6 drop and reduce immune system power. However, a nap can bring those levels back to normal. Researchers have also found increased levels of norepinephrine, the "fight or flight" molecule, after reduced night-time sleep. Sustained high levels of norepinephrine, associated with the stress response, have harmful effects on blood glucose balance and cardiovascular health. Napping brought the norepinephrine levels back within their normal range. Similarly, since one in three adults in the US have high blood pressure, it is welcome news that a daily nap can bring that down as effectively as medications and other lifestyle changes.

One definition of a good citizen is someone who acts in ways that benefit society; that is, the benefits are a consequence of the citizen's actions. A good citizen votes and pays taxes, for example, because both actions help to create stable and prosperous societies. In contrast, a bad citizen is one who breaks the law, to the extent that breaking the law harms other people. In this view, someone who speeds would be a bad citizen because speeding increases the likelihood of causing a crash and harming others, but someone who commits a "victimless crime," such as smoking marijuana, would not be a bad citizen because they would not be harming anyone else. According to normative political science, a person should behave in ways that benefit society and do not harm it, and individuals should strive to be good citizens.

B) Scene viewing

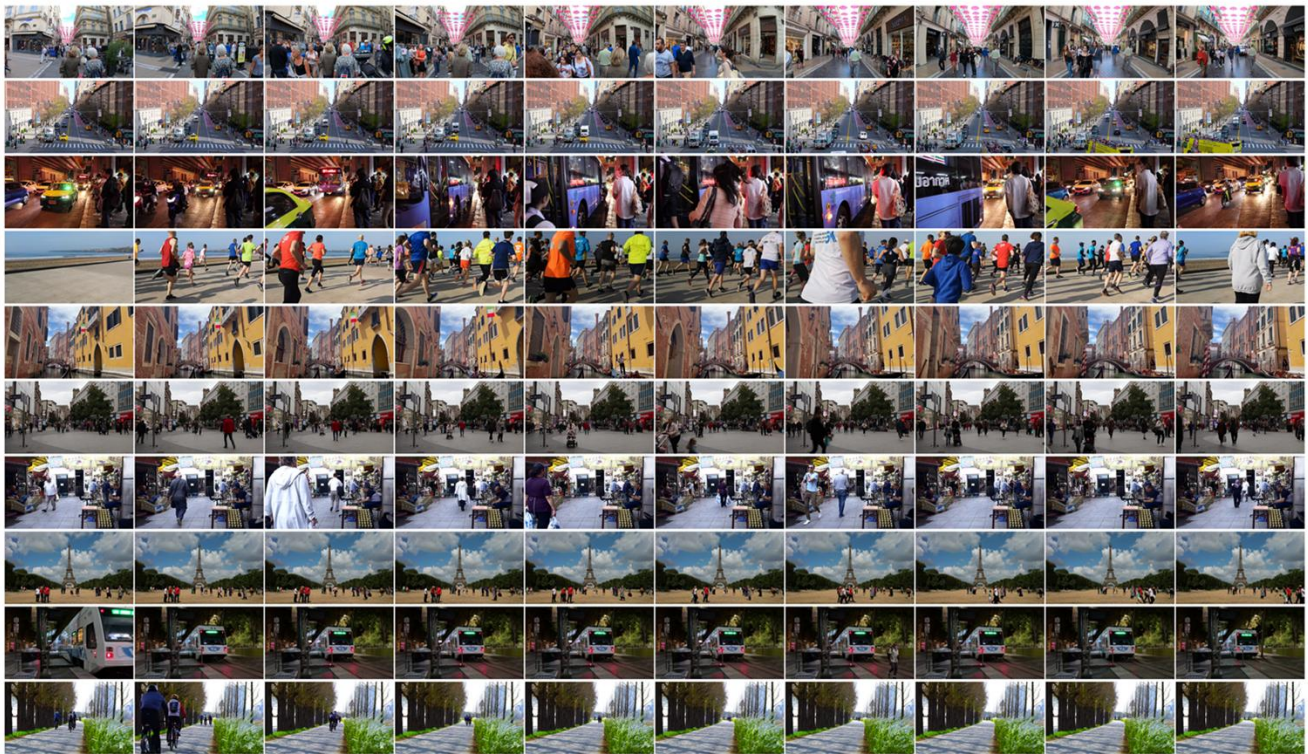


Figure 1. A sample of the stimuli used in the reading task (a) and the scene viewing task (b).

Panel b shows video frames taken from a subset of the videos. The frames were selected at 10 evenly spaced time intervals based on the overall length of the video. *Note:* all videos are under a free-to-use public domain license from [Pexels.com](https://www.pexels.com/) and [Pixabay.com](https://pixabay.com/).

Sound stimuli. The standard sound was a 400 Hz sine-wave tone, and the novel sounds consisted of 100 different environmental sounds (e.g. a knock, a fire alarm, a whistle, drilling, glass breaking). Eighty novel sounds were taken from Andrés et al. (2006), who adapted them from Escera et al. (1998), and 20 sounds were taken from Vasilev (2021). All sounds were 150 ms long and were sampled at 44.1 kHz, 16-bit, stereo. Due to their short duration, the novel sounds were usually not easily identifiable.

Apparatus

Eye-movements were recorded with an Eyelink 1000 (SR Research, Ontario, Canada) in a Tower mount set-up at a sampling frequency of 1000 Hz. This system has an average spatial accuracy of 0.25-0.5°. Participants' heads were stabilised with a chin-and-forehead rest. Viewing was binocular, but only the right eye was recorded¹. Testing was done in a small, windowless lab cubicle, illuminated with an overhead LED lamp.

The experiment was programmed in Matlab 2021b (MathWorks, 2021) using the Psychophysics Toolbox v.3.15 (Brainard, 1997; Cornelissen et al., 2002; Kleiner et al., 2007; Pelli, 1997). The visual stimuli were displayed on a 24.5" Alienware 25 LCD monitor (resolution: 1920 × 1080; refresh rate: 244 Hz). The width of the screen was 54 cm, and the eye-to-screen distance was 62 cm, with each pixel subtending 0.024 degrees of visual angle horizontally. The sounds were played binaurally at 65 ± 1.5 dB(A) via a Creative Sound Blaster Z sound card (SB1500), connected to Bose QuietComfort 25 Noise-Cancelling headphones. The audio playback had a true latency of 25 ms (measured with the Black Box ToolKit v2, Sheffield, UK).

¹ The left eye was recorded for 5 participants due to reflections with wearing glasses/contact lenses on the right eye.

The Matlab script took that delay into account to ensure that sounds were presented 120 ms after fixation onset (see below).

Procedure

Participants provided written informed consent, after which a 9-point calibration and validation routine was performed. Calibration accuracy better than 0.5° was maintained throughout the experiment. A fixation check was presented before each trial, and participants were recalibrated whenever necessary. Each task block started with two practice items (during which no sounds were played), followed by the experimental trials. Participants read the passage at their own pace and pressed the left button of the mouse to terminate the trial (trials ended automatically after 1 minute if there was no response). In the scene viewing task, the trial ended after the video finished playing. Participants were instructed to focus on the task at hand and ignore any sounds they may hear.

In each trial, saccades were parsed in real time from the eye position data using Eyelink's saccade detection algorithm. This method uses a threshold-based algorithm in which a saccade is detected when the eye's motion exceeds a specific acceleration and velocity threshold. In this experiment, the acceleration threshold was $9500^\circ/\text{s}$, and the velocity threshold was $35^\circ/\text{s}$. Because the algorithm requires 35 samples for its calculation, saccades were detected only 35 ms after they had started. The software counted the saccades from the start of the trial and presented a sound on every 5th fixation (Graupner et al., 2007). The sounds were presented 120 ms after the onset of fixation (Vasilev et al., 2021), taking into account the saccade detection delay and the sound latency delay². After the end of the trial, a reading comprehension question or a memory

² This is comparable to the delay of 100 ms used by Graupner et al. (2007).

question (in the scene viewing task) was presented. Participants were offered a break halfway through the experiment. The study took about 30-40 minutes to complete.

Data Analysis

Data were analysed using Linear Mixed Models using the “lme4” v.1.1-34 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages in R 4.42 (R Core Team, 2024). Participants and items were added as random intercepts (Baayen et al., 2008). Additionally, we attempted to add Task and Sound as random slopes (Barr et al., 2013). If the models failed to converge, the slopes were removed one by one until convergence was achieved. Sum contrasts coding was used for the task (reading= 1; video = -1) and sound (novel= 1; standard= -1) conditions. Fixation durations were log-transformed prior to analysis.

We also calculated supplementary Bayes Factors with the brms package v.2.20.0 (Bürkner, 2018), using the same model structure as the frequentist models. We used Normal (0, 5) priors for the intercept and Normal (0, 0.08) priors for the slopes, which roughly corresponds to the expected effect size on the log scale (Vasilev et al., 2023). Posterior sampling was done with 10 chains (5000 iterations each, including 1000 for warm-up). Bayes Factors (BF_{10}) were calculated using the Savage-Dickey density ratio method (Dickey & Lientz, 1970; Morey et al., 2011). Empirical effect sizes in Cohen’s d (Cohen, 1988) are also reported.

Results

Out of the 32000 sounds played in the study, 29762 (93%) were played accurately during a fixation³. Fixations that were preceded or followed by a blink ($n= 3722$; 11.63%) were discarded. Additionally, fixations shorter than 50 ms or longer than 2000 ms were discarded as outliers ($n= 87$; 0.27%). This left 25953 fixations (81.10% of the data) for analysis. The average accuracy was 84.6% ($SD= 36.1\%$) in the reading task and 87.4% ($SD= 33.2\%$) in the scene viewing task. There were no significant differences in accuracy between the two tasks ($b= -0.3210$, $SE= 0.2017$, $z= -1.592$, $p= 0.11$).

The fixation duration data are visualised in Figure 2, and the statistical results are presented in Table 1. There was a main effect of Task, with scene viewing ($M= 331$ ms; $SD= 201$ ms) resulting in longer fixation durations compared to reading ($M= 233$ ms; $SD= 96.1$ ms). Additionally, there was a significant main effect of Sound: novel sounds ($M= 291$ ms, $SD=168$ ms) led to higher fixation durations compared to standard sounds ($M= 279$; $SD=163$ ms) across both tasks. The interaction between Task and Sound was not significant but was close to the .05 threshold. In general, the effect of novel sounds was less pronounced in the scene viewing task compared to the reading task. As Figure 3 shows, the novelty distraction effect was more variable across the different items in the dynamic scene viewing task. The Bayes Factors very strongly favoured the alternative hypothesis for the main effects of Task and Sound (indicating the presence of a difference). However, the Bayes Factor for the interaction favoured the null hypothesis of no difference, suggesting that the novelty distraction effect was similar across both tasks. In summary, there was evidence for general inhibition across both tasks. While the degree of

³ Almost all the remaining fixations ($n= 2233$; 6.97%) were played in a saccade and several ($n= 5$; 0.01%) were played just when the trial was ending. Sounds were played in a saccade when the fixation was terminated early, and the sound appeared in the following saccade.

inhibition was numerically weaker in the dynamic scene viewing task, the statistical models did not show a reliable difference between the two tasks.

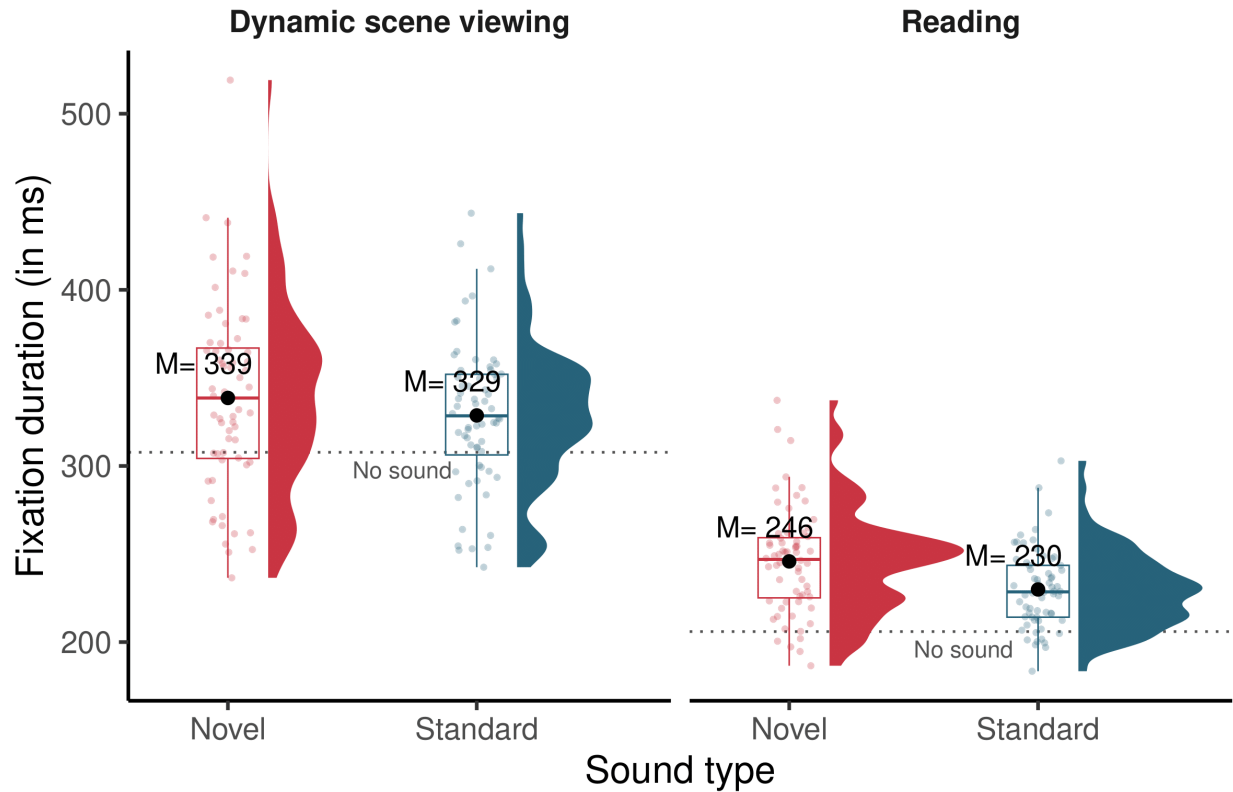


Figure 2. Mean fixation durations (aggregated over participants) in the two Sound and Task conditions. Dots show by-participants means. The larger black dots indicate the condition mean aggregated across all participants. Horizontal dotted lines indicate the mean duration of fixations occurring between sounds.

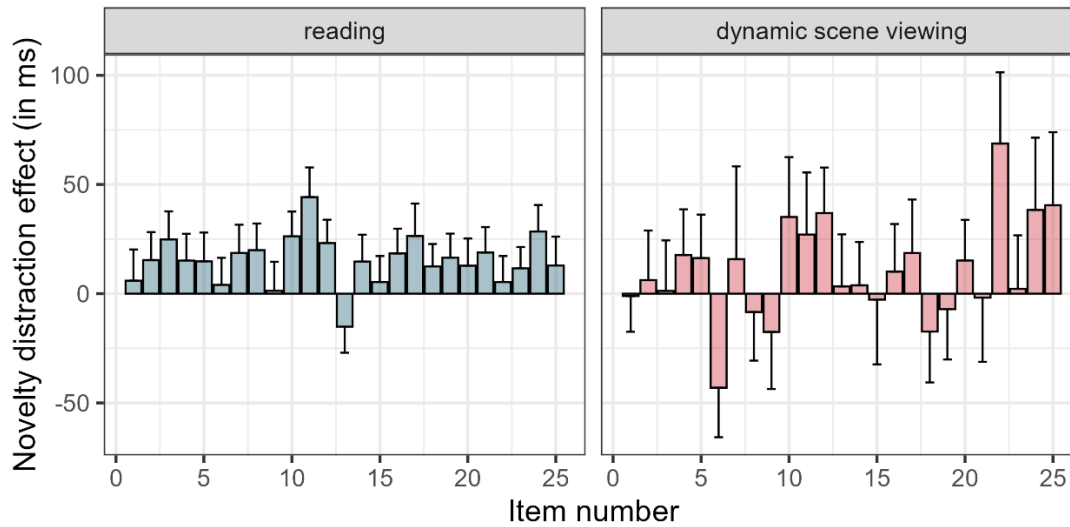


Figure 3. Novelty distraction effect size (Novel- Standard) in fixation durations, averaged for each item across both tasks. Error bars indicate ± 1 SE.

Table 1

Results of the Linear Mixed Effect Model for Fixation Durations

<i>Predictors</i>	log(Fixation duration)				<i>BF</i> ₁₀
	<i>Estimate</i>	<i>Std. Error</i>	<i>Statistic</i>	<i>p</i>	
(Intercept)	5.546	0.014	402.463	<0.001	
Task (reading vs. dynamic scene viewing)	-0.146	0.010	-14.306	<0.001	2.377e+16
Sound (novel vs. standard)	0.023	0.003	7.044	<0.001	5.749e+14
Task \times Sound	0.006	0.003	1.895	0.058	0.234
Random Effects					
σ^2	0.173				
τ_{00} subjects (intercept)	0.008				
τ_{00} items (intercept)	0.003				
τ_{11} subjects (random slope for Task)	0.002				
ρ_{01} subjects (slope-intercept correlation)	-0.315				
ICC (inter-class correlation)	0.068				

N_{subjects}	64
N_{items}	50
Observations	25953
Marginal R^2 / Conditional R^2	0.109 / 0.169

Note: model summary generated by the sjPlot package v.2.8.17 (Lüdtke, 2024). BF_{10} ,

Bayes factor comparing the alternative hypothesis to the null hypothesis. While Bayes Factors < 0.33 indicate evidence for the null hypothesis, Bayes factors > 3 indicate evidence for the alternative hypothesis.

Influence of smooth pursuit eye-movements. The eye-movement event parser of the EyeLink system does not include a specific detector for smooth pursuit movements. Because smooth pursuit movements typically have lower velocity and acceleration than saccades, fixations recorded during the dynamic scene viewing task may include some smooth pursuit segments. Consequently, some sounds may have been played during smooth pursuit rather than fixation events. To assess whether this affected the results, we performed a post-hoc analysis by parsing the raw gaze data through Startsev et al.'s (2019) smooth pursuit detection algorithm, trained on the GazeCom dataset (Dorr et al., 2010). Removing fixations that contained sequential smooth pursuit samples for at least half of their duration (11.1% of the data) did not change the main results above (see Table S1 in the Supplementary Materials)⁴. A further model was fitted on the scene viewing data alone, by adding the proportion of samples identified as belonging to smooth-pursuit events within each fixation as a covariate (see Table S2 in the Supplementary Materials). The results showed that there was a numeric trend for the novelty distraction effect to become smaller when more smooth-pursuit samples were present within a fixation, but the

⁴ A more conservative approach of removing fixations with *any* smooth pursuit samples (24.5% data loss) also led to the same conclusion.

interaction effect was not significant ($p = 0.192$). Therefore, while smooth pursuits may account for some of the variability in the novelty inhibition effect, they do not fully explain it.

Discussion

The present study tested whether unexpected sounds lead to a general inhibition of eye-movements during two visual-cognitive tasks: passage reading and dynamic scene viewing. Building on previous work from static scene viewing (Graupner et al., 2007), we presented sounds on every 5th fixation to examine oculomotor inhibition in two complex and ecologically valid tasks. There were three main findings from the experiment. First, fixation durations were considerably longer during dynamic scene viewing compared to passage reading, even after excluding fixations that contained smooth pursuits. This aligns with previous findings that fixation durations are longer when viewing static scenes than when reading text (e.g., Nuthmann & Henderson, 2012) and longer when viewing dynamic compared to static scenes (e.g., Dorr et al., 2010; Smith & Mital, 2013). Second, unexpected novel sounds led to longer fixation durations compared to standard sounds in both tasks, also replicating previous evidence of oculomotor inhibition (e.g., Kadosh & Bonne, 2022; Rettie et al., 2024; Valsecchi & Turatto, 2009; Vasilev et al., 2023, 2025). Third, there was generally weak evidence that the degree of inhibition differed across the two tasks. While the effect was numerically weaker in the dynamic scene viewing task compared to the reading task, the Bayes factor nevertheless suggested that there is no difference between them.

The main effect of novel sounds on fixation durations supports the view that perceptual surprise triggers a general inhibition of motor actions that may facilitate the processing of unexpected events and thus help observers select an appropriate response (Diesburg & Wessel,

2021; Wessel, 2017, 2018). The effect was observed both during reading, which is a well-practiced and highly stereotypical task, and during dynamic scene viewing, where oculomotor behaviour is more variable and prone to bottom-up influences from the scene (Dorr et al., 2010). Not surprisingly, the dynamic scene viewing task showed more variability in the sound distraction effect across items (see Figure 3). Future research is required to pinpoint the perceptual or cognitive processes that modulate oculomotor inhibition in dynamic scene viewing. The more variable and slightly weaker inhibition effect in the scene viewing task suggests that motor inhibition is not completely indiscriminate of the task's characteristics. Rather, task demands or stimulus properties could modulate the strength of this response.

Dynamic scenes contain strong exogenous cues (e.g., motion and flicker), which are key predictors of gaze location (Carmi & Itti, 2006; Itti, 2005; Mital et al., 2011). This typically results in greater attentional synchrony in dynamic compared to static viewing tasks (Dorr et al., 2010; Smith & Mital, 2013). Therefore, because dynamic scene viewing involves greater bottom-up oculomotor control compared to reading, it may make observers less susceptible to other bottom-up stimuli (such as irrelevant novel sounds), which are competing for attentional resources. This would suggest that while the motor system is broadly susceptible to inhibition from unexpected sensory input (Wessel & Aron, 2017), the degree to which this occurs may partly reflect the task dynamics.

In summary, the present study used an existing methodology (Graupner et al., 2007) in a new way to study saccadic inhibition by unexpected sounds in two ecologically valid tasks. The results suggest that eye-movements are broadly susceptible to novelty inhibition, supporting the idea that perceptual surprise can trigger a global inhibition of motor actions (Wessel, 2017, 2018; Wessel & Aron, 2013). At the same time, oculomotor inhibition was more variable and

numerically weaker in the dynamic scene viewing task, even if no reliable task differences were found. More research is needed to understand if bottom-up visual processes in the task may modulate oculomotor responses to perceptually salient sounds.

Funding:

This work was funded by a small grant from the Experimental Psychology Society (UK) awarded to MRV, as well as a research grant PID2020-114117GB-I00 awarded to F.B.R.P. from MICIU/AEI /10.13039/501100011033. The study was conducted while M.R.V. was at Bournemouth University.

Data availability

Data, analysis scripts and materials from this study are available at: <https://osf.io/3wjsp/>

Conflict of Interests: None

References

- Andrés, P., Parmentier, F. B. R., & Escera, C. (2006). The effect of age on involuntary capture of attention by irrelevant sounds: A test of the frontal hypothesis of aging. *Neuropsychologia*, 44(12), 2564–2568. <https://doi.org/10.1016/j.neuropsychologia.2006.05.005>
- Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Berti, S. (2008). Cognitive control after distraction: Event-related brain potentials (ERPs) dissociate between different processes of attentional allocation. *Psychophysiology*, 45(4), 608–620. <https://doi.org/10.1111/j.1469-8986.2008.00660.x>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Carmi, R., & Itti, L. (2006). The role of memory in guiding attention during natural vision. *Journal of Vision*, 6(9), 4. <https://doi.org/10.1167/6.9.4>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617. <https://doi.org/10.3758/BF03195489>
- Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov Chain. *The Annals of Mathematical Statistics*, 41(1), 214–226. <https://www.jstor.org/stable/2239734>
- Diesburg, D. A., & Wessel, J. R. (2021). The Pause-then-Cancel model of human action-stopping: Theoretical considerations and empirical evidence. *Neuroscience and Biobehavioral Reviews*, 129(April), 17–34. <https://doi.org/10.1016/j.neubiorev.2021.07.019>
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10), 1–17. <https://doi.org/10.1167/10.10.28>
- Dutra, I. C., Waller, D. A., & Wessel, J. R. (2018). Perceptual surprise improves action stopping by nonselectively suppressing motor activity via a neural mechanism for motor inhibition. *Journal of Neuroscience*, 38(6), 1482–1492. <https://doi.org/10.1523/JNEUROSCI.3091-17.2017>
- Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, 10(5), 590–604. <https://doi.org/10.1162/089892998562997>
- Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain

- potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4), 355–373. [https://doi.org/10.1016/S0149-7634\(01\)00019-7](https://doi.org/10.1016/S0149-7634(01)00019-7)
- Furubayashi, T., Ugawa, Y., Terao, Y., Hanajima, R., Sakai, K., Machii, K., Mochizuki, H., Shiio, Y., Uesugi, H., Enomoto, H., & Kanazawa, I. (2000). The human hand motor area is transiently suppressed by an unexpected auditory stimulus. *Clinical Neurophysiology*, 111(1), 178–183. [https://doi.org/10.1016/S1388-2457\(99\)00200-X](https://doi.org/10.1016/S1388-2457(99)00200-X)
- García-López, E., Parmentier, F. B. R., Kaiser, J., & Böer, N. T. (2024). *Distraction by unexpected sounds : comparing response repetition and response switching*. October, 1–12. <https://doi.org/10.3389/fpsyg.2024.1451008>
- Goettker, A., Agtzigdis, I., Braun, D. I., Dorr, M., & Gegenfurtner, K. R. (2020). From Gaussian blobs to naturalistic videos: Comparison of oculomotor behavior across different stimulus complexities. *Journal of Vision*, 20(8), 26. <https://doi.org/10.1167/jov.20.8.26>
- Goettker, A., & Gegenfurtner, K. R. (2021). A change in perspective: The interaction of saccadic and pursuit eye movements in oculomotor control and perception. *Vision Research*, 188, 283–296. <https://doi.org/10.1016/j.visres.2021.08.004>
- Graupner, S. T., Velichkovsky, B. M., Pannasch, S., & Marx, J. (2007). Surprise, surprise: Two distinct components in the visually evoked distractor effect. *Psychophysiology*, 44(2), 251–261. <https://doi.org/10.1111/j.1469-8986.2007.00504.x>
- Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Horstmann, G., & Herwig, A. (2015). Surprise attracts the eyes and binds the gaze. *Psychonomic*

- Bulletin and Review*, 22(3), 743–749. <https://doi.org/10.3758/s13423-014-0723-1>
- Huey, E. B. (1908). *The psychology and pedagogy of reading: With a review of the history of reading and writing and of methods, texts, and hygiene in reading*. MacMillan.
- Iacullo, C., Diesburg, D. A., & Wessel, J. R. (2020). Non-selective inhibition of the motor system following unexpected and expected infrequent events. *Experimental Brain Research*. <https://doi.org/10.1101/2020.03.25.008789>
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6), 1093–1123. <https://doi.org/10.1080/13506280444000661>
- Kadosh, O., & Bonne, Y. S. (2022). Involuntary oculomotor inhibition markers of saliency and deviance in response to auditory sequences. *Journal of Vision*, 22(5), 1–19. <https://doi.org/10.1167/jov.22.5.8>
- King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS ONE*, 9(1), e85791. <https://doi.org/10.1371/journal.pone.0085791>
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *30th European Conference on Visual Perception*. https://pure.mpg.de/pubman/item/item_1790332_4/component/file_3136265/ECVP-2007-Kleiner.pdf
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13), 1457–1483. <https://doi.org/10.1016/j.visres.2010.12.014>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in

- linear mixed effects models. *Journal of Statistical Software*, 82(13).
<https://doi.org/10.18637/jss.v082.i13>
- Lüdtke, D. (2024). *sjPlot: Data visualization for statistics in social science* (2.8.17).
<https://cran.r-project.org/package=sjPlot>
- Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Hubel, D. H. (2009). Microsaccades: a neurophysiological analysis. *Trends in Neurosciences*, 32(9), 463–475.
<https://doi.org/10.1016/j.tins.2009.05.006>
- MathWorks. (2021). *Matlab R2021b [Computer software]*.
- Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, 3(1), 5–24.
<https://doi.org/10.1007/s12559-010-9074-z>
- Morey, R. D., Rouder, J. N., Pratte, M. S., & Speckman, P. L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology*, 55(5), 368–378. <https://doi.org/10.1016/j.jmp.2011.06.004>
- Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory–cognitive core of cognitive processes. *Brain Research Reviews*, 64(1), 123–136. <https://doi.org/10.1016/j.brainresrev.2010.03.001>
- Novembre, G., Pawar, V. M., Bufacchi, R. J., Kilintari, M., Srinivasan, M., Rothwell, J. C., Haggard, P., & Iannetti, G. D. (2018). Saliency detection as a reactive process: Unexpected sensory events evoke corticomuscular coupling. *The Journal of Neuroscience*, 38(9), 2385–2397. <https://doi.org/10.1523/JNEUROSCI.2474-17.2017>
- Nuthmann, A., & Canas-Bajo, T. (2022). Visual search in naturalistic scenes from foveal to

- peripheral vision: A comparison between dynamic and static displays. *Journal of Vision*, 22(1), 10. <https://doi.org/10.1167/jov.22.1.10>
- Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, 20(4–5), 457–494. <https://doi.org/10.1080/13506285.2012.670142>
- Parmentier, F. B. R., Vasilev, M. R., & Andrés, P. (2019). Surprise as an explanation to auditory novelty distraction and post-error slowing. *Journal of Experimental Psychology: General*, 148(1), 192–200. <https://doi.org/10.1037/xge0000497>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- R Core Team. (2024). *R: A language and environment for statistical computing* (4.31). R Foundation for Statistical Computing. <http://www.r-project.org/>
- Rettie, L., Marsh, J. E., Liversedge, S. P., Wang, M., & Degno, F. (2024). Auditory distraction during reading: Investigating the effects of background sounds on parafoveal processing. *Quarterly Journal of Experimental Psychology*.
- Roberts, J. A., Wallis, G., & Breakspear, M. (2013). Fixational eye movements during viewing of dynamic natural scenes. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00797>
- Rolfs, M., Kliegl, R., & Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision*, 8(11), 5–5. <https://doi.org/10.1167/8.11.5>
- Sabu, S., Parmentier, F. B. R., & Horváth, J. (2024). Involuntary motor responses are elicited

- both by rare sounds and rare pitch changes. *Scientific Reports*, 14(1), 20235.
<https://doi.org/10.1038/s41598-024-70776-x>
- Schröger, E., & Wolff, C. (1998). Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. *Cognitive Brain Research*, 7(1), 71–87.
[https://doi.org/10.1016/S0926-6410\(98\)00013-5](https://doi.org/10.1016/S0926-6410(98)00013-5)
- Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 16–16.
<https://doi.org/10.1167/13.8.16>
- Startsev, M., Agtzigdis, I., & Dorr, M. (2019). 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods*, 51(2), 556–572.
<https://doi.org/10.3758/s13428-018-1144-2>
- Stephenson, A. C., Eimontaite, I., Caleb-Solly, P., Morgan, P. L., Khatun, T., Davis, J., & Alford, C. (2020). Effects of an unexpected and expected event on older adults' autonomic arousal and eye fixations during autonomous driving. *Frontiers in Psychology*, 11.
<https://doi.org/10.3389/fpsyg.2020.571961>
- Valsecchi, M., & Turatto, M. (2009). Microsaccadic responses in a bimodal oddball task. *Psychological Research*, 73(1), 23–33. <https://doi.org/10.1007/s00426-008-0142-x>
- Vasilev, M. R. (2021). *Collection of novel sounds for auditory distraction experiments*.
<https://doi.org/10.17605/OSF.IO/PH7TE>
- Vasilev, M. R., Lowman, M., Bills, K., Parmentier, F. B. R., & Kirkby, J. A. (2023). Unexpected sounds inhibit the movement of the eyes during reading and letter scanning. *Psychophysiology*, 60(12), 1–19. <https://doi.org/10.1111/psyp.14389>

- Vasilev, M. R., Ozkan, Z. G., Kirkby, J. A., Nuthmann, A., & Parmentier, F. B. R. (2025). Unexpected sounds induce a rapid inhibition of eye-movement responses. *Psychophysiology*, 62(1), 1–19. <https://doi.org/10.1111/psyp.14728>
- Vasilev, M. R., Parmentier, F. B., Angele, B., & Kirkby, J. A. (2019). Distraction by deviant sounds during reading: An eye-movement study. *Quarterly Journal of Experimental Psychology*, 72(7), 1863–1875. <https://doi.org/10.1177/1747021818820816>
- Vasilev, M. R., Parmentier, F. B., & Kirkby, J. A. (2021). Distraction by auditory novelty during reading: Evidence for disruption in saccade planning, but not saccade execution. *Quarterly Journal of Experimental Psychology*, 74(5), 826–842. <https://doi.org/10.1177/1747021820982267>
- Wessel, J. R. (2017). Perceptual surprise aides inhibitory motor control. *Journal of Experimental Psychology: Human Perception and Performance*, 43(9), 1585–1593. <https://doi.org/10.1037/xhp0000452>
- Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, 55(3), 1–21. <https://doi.org/10.1111/psyp.13041>
- Wessel, J. R., & Aron, A. R. (2013). Unexpected events induce motor slowing via a brain mechanism for action-stopping with global suppressive effects. *Journal of Neuroscience*, 33(47), 18481–18491. <https://doi.org/10.1523/JNEUROSCI.3456-13.2013>
- Wessel, J. R., & Aron, A. R. (2017). On the Globality of Motor Suppression: Unexpected Events and Their Influence on Behavior and Cognition. *Neuron*, 93(2), 259–280. <https://doi.org/10.1016/j.neuron.2016.12.013>
- Widmann, A., Engbert, R., & Schröger, E. (2014). Microsaccadic responses indicate fast

categorization of sounds: A novel approach to study auditory cognition. *Journal of Neuroscience*, 34(33), 11152–11158. <https://doi.org/10.1523/JNEUROSCI.1568-14.2014>

Zhao, S., Yum, N. W., Benjamin, L., Benhamou, E., Yoneya, M., Furukawa, S., Dick, F., Slaney, M., & Chait, M. (2019). Rapid ocular responses are modulated by bottom-up-driven auditory salience. *The Journal of Neuroscience*, 39(39), 7703–7714. <https://doi.org/10.1523/JNEUROSCI.0776-19.2019>

Supplemental Materials

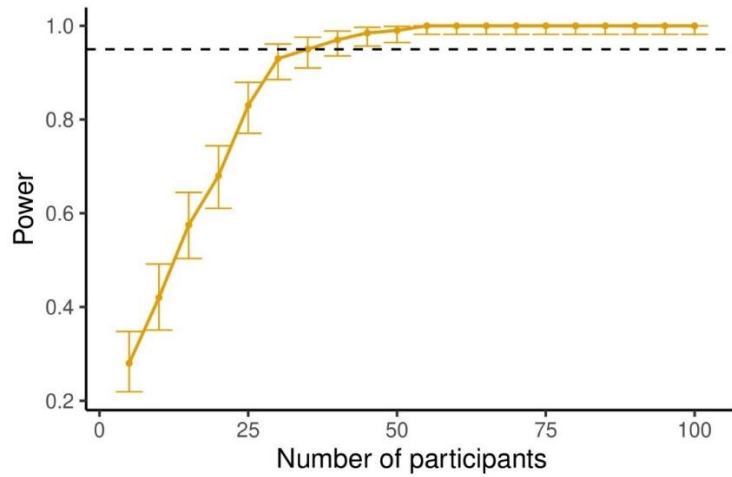


Figure S1. Statistical power simulations based on data from (Vasilev et al., 2021). Horizontal dotted line shows power at 95%. In the simulations, we assumed an $\alpha=0.05$ and effect size that is 75% of the expected effect size ($d \approx 0.16$). The LMM was fitted with random intercepts for subjects and items, and a random slope for sound for subjects. The simulations are based on 200 runs per each increment of 5 subjects.

Main Results after Excluding Fixations with Smooth-pursuit Events

Table S1

Results of the Linear Mixed Effect Model for Fixation Duration after Excluding Fixations Containing Smooth-pursuit Events

<i>Predictors</i>	log(Fixation duration)				<i>BF₁₀</i>
	<i>Estimate</i>	<i>Std. Error</i>	<i>Statistic</i>	<i>p</i>	
(Intercept)	5.554	0.014	387.672	<0.001	
Task (reading vs. dynamic scene viewing)	0.154	0.011	14.467	<0.001	6.474e+15
Sound (novel vs. standard)	0.024	0.003	7.027	<0.001	4.192e+15
Task × Sound	-0.005	0.003	-1.435	0.151	0.12886
Random Effects					
σ^2	0.167				
τ_{00} subjects (intercept)	0.009				
τ_{00} items (intercept)	0.003				
τ_{11} subjects (random slope for Task)	0.003				
ρ_{01} subjects (slope-intercept correlation)	0.441				
ICC (inter-class corelation)	0.076				
N _{subjects}	64				
N _{items}	50				
Observations	23078				
Marginal R ² / Conditional R ²	0.119 / 0.186				

Note: model summary generated by the sjPlot package v.2.8.17 (Lüdtke, 2024). BF_{10} , Bayes factor comparing the alternative hypothesis to the null hypothesis. While Bayes Factors < 0.33 indicate evidence for the null hypothesis, Bayes factors >3 indicate evidence for the alternative hypothesis.

Table S2

Supplementary Model Adding the Percentage of Smooth-Pursuit Samples within a Fixation as a Covariate for the Dynamic Scene Viewing Data

log(Fixation duration)				
<i>Predictors</i>	<i>Estimate</i>	<i>std. Error</i>	<i>Statistic</i>	<i>p</i>
(Intercept)	5.691	0.021	268.582	<0.001
Sound (novel vs. standard)	0.016	0.006	2.802	0.005
SP perc.	0.001	< 0.001	5.808	<0.001
Sound × SP perc.	< -0.001	< 0.001	-1.305	0.192
Random Effects				
σ^2	0.228			
τ_{00} subjects	0.013			
τ_{00} items	0.006			
τ_{11} subjects (random slope for Sound)	0.000			
ρ_{01} subjects (slope-intercept correlation)	0.307			
ICC (inter-class correlation)	0.072			
N_{subjects}	64			
N_{items}	25			
Observations	12751			
Marginal R^2 / Conditional R^2	0.007 / 0.079			

Note: SP perc: percentage of samples within each fixation which were identified as belonging to smooth pursuit events. Note that *SP perc* was centred in the model at 0.

Model summary generated by the sjPlot package v.2.8.17 (Lüdtke, 2024).